



Planning optimization in prostate and head-and-neck radiation therapy

Pengcheng Zhang

► To cite this version:

Pengcheng Zhang. Planning optimization in prostate and head-and-neck radiation therapy. Other. Université de Rennes; Southeast university (Nanjing, China), 2014. Chinese. NNT : 2014REN1S042 . tel-01077977

HAL Id: tel-01077977

<https://theses.hal.science/tel-01077977>

Submitted on 27 Oct 2014

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



THÈSE / UNIVERSITÉ DE RENNES 1
sous le sceau de l'Université Européenne de Bretagne

En Cotutelle Internationale avec
Southeast University, Nankin, Chine

pour le grade de
DOCTEUR DE L'UNIVERSITÉ DE RENNES 1
Mention : Traitement du Signal et Télécommunications

Ecole doctorale Matisse

présentée par

Pengcheng ZHANG

Préparée à l'unité de recherche LTSI – INSERM U1099
Laboratoire Traitement du Signal et de l'Image
Université de Rennes 1, France
Laboratory of Image Science & Technology (LIST)
Southeast University, Nanjing, Chine

**Optimisation de
la planification
en radiothérapie
prostatique et
ORL**

**Soutenance à Nankin, Chine
le 02/07/2014**

devant le jury composé de :

Pascal FRANCOIS

Physicien Médical, CHU de Poitiers / *rapporteur*

Zhiguo GUI

Professeur, North University of China / *rapporteur*

Baosheng LI

PUPH, Shandong Cancer Hospital and Institute / *examinateur*

Limin LUO

Professeur, Southeast University / *examinateur*

Huazhong SHU

Professeur, Southeast University / *directeur de thèse*

Renaud DE CREVOISIER

PUPH, Université de Rennes 1 / *directeur de thèse*

Antoine SIMON

MCU, Université de Rennes 1 / *co-directeur de thèse*

Résumé en Français

Optimisation de la planification en radiothérapie prostatique et ORL

Le cancer correspond à une prolifération cellulaire anormalement importante, la tumeur, au sein d'un tissu sain dont la survie est alors menacée. Selon les dernières estimations du Centre international de recherche sur le cancer (CIRC) et de l'Organisation mondiale de la Santé (OMS), environ dix millions de nouveaux cas de cancer sont diagnostiqués chaque année, avec un peu plus de la moitié des cas dans les pays en développement. En 2015, ce nombre devrait atteindre environ 15 millions de cas, dont les deux tiers dans les pays en développement. Le traitement du cancer en Chine est ainsi une tâche difficile, le taux de mortalité par cancer ayant augmenté de près de 30% au cours des 20 dernières années. En 2008, les nouveaux cas de cancer ont atteint 2,82 millions et les prévisions donnent un nombre de 21 millions en 2030.

La radiothérapie est devenue l'un des trois principaux traitements du cancer, avec la chimiothérapie et la chirurgie. Elle repose sur l'utilisation de radiations, généralement des photons, pour détruire les cellules cancéreuses en bloquant leur capacité à se multiplier. L'objectif de la radiothérapie est de délivrer une dose prescrite à la tumeur tout en minimisant l'irradiation des tissus sains environnants. Un traitement par radiothérapie repose sur deux étapes principales : (i) la planification du traitement, durant laquelle les caractéristiques de l'irradiation sont optimisées suivant l'anatomie du patient telle qu'elle est représentée par une image scanner ; (ii) la délivrance du traitement fractionnée en une série de séances (par exemple 40 fractions pour un traitement du cancer de la prostate). Si les nombreuses avancées technologiques des dernières années, que ce soit en imagerie (scanner, IRM et TEP), pour le calcul de dose (superposition 3D, Monte-Carlo) ou pour la délivrance du traitement (IMRT, SRS, IGRT), ont permis d'améliorer considérablement la précision du traitement, un certain nombre de problématiques restent ouvertes. Ainsi, lors de la planification, les algorithmes de calcul de dose peuvent encore souffrir d'imprécisions. La prise en compte de paramètres radio-biologiques lors de cette planification reste aussi à améliorer. De même, l'impact des incertitudes géométriques lors de la délivrance du traitement doit être compensé. Enfin, dans une stratégie de radiothérapie adaptative, permettant de prendre en compte des modifications anatomiques en cours de traitement par des replanifications, les scénarii optimaux restent à identifier. Ce travail de thèse s'attache à proposer des solutions à ces différents enjeux.

Lors de la planification, la balistique de traitement est optimisée suivant l'anatomie du patient telle qu'elle est représentée par le scanner de planification. Cette étape repose notamment sur la simulation, pour une balistique de traitement donnée, de la matrice de dose reçue par les différents organes. Différentes méthodes sont utilisées pour cette simulation dans un objectif de compromis entre précision et rapidité de calcul. L'approche la plus fréquemment utilisée est celle du *Pencil Beam*, reposant sur la convolution/superposition de faisceaux élémentaires. Cette approche a cependant certaines limites : (i) l'inclinaison des faisceaux élémentaires est négligée ; (ii) la correction des hétérogénéités de tissus est approximative ; (iii) ces approches restent coûteuses en temps de calcul. Dans ce travail, l'approche classique du *Pencil Beam* a été modifiée en considérant un système de coordonnées sphériques adapté à la géométrie des faisceaux, en modifiant le mode de correction des hétérogénéités et en accélérant le calcul grâce au calcul des convolutions par la transformée de Fourier rapide. L'approche proposée a été comparée à la méthode AAA en considérant le calcul de Monte-Carlo comme référence. Différents fantômes numériques (eau, eau et insert de poumon, eau et insert d'os, eau et inserts de poumon et d'os) ont été utilisés. La méthode proposée a ainsi montré une meilleure précision que la méthode AAA, avec cependant des erreurs importantes aux niveaux des interfaces. Par ailleurs, le calcul par transformée de Fourier permet d'accélérer le calcul d'un facteur 40, au prix toutefois d'une dégradation de la précision des résultats. Il serait ainsi intéressant, dans de futurs travaux, de combiner les deux approches : le calcul par transformée de Fourier lorsqu'une estimation grossière de la dose est suffisante, puis par la méthode par superposition pour affiner le résultat de l'optimisation.

Toujours lors de la planification, une fois que la matrice de dose a été simulée, une fonction de coût est considérée de façon à pouvoir comparer quantitativement différentes balistiques. Cette fonction de coût repose généralement sur des contraintes dosimétriques simples issues notamment de l'histogramme dose-volume, et ne prend pas en compte la complexité de la réponse biologique de la tumeur ou des tissus sains. Ceci pourrait être corrigé par l'incorporation de modèles radio-biologiques tels que les modèles NTCP (probabilité de toxicité des tissus sains). Cependant, l'incorporation de tels modèles est compliquée par leur non-linéarité et non-convexité, rendant ainsi l'optimisation du plan de traitement beaucoup plus complexe. Dans ce travail, un critère radio-biologique correspondant à l'équivalent convexe du modèle NTCP a été utilisé. Une fonction de coût incorporant à la fois ce critère radio-biologique (pour limiter l'irradiation des tissus sains) et des critères physiques (pour assurer la couverture de la cible) a alors été proposée pour constituer un modèle hybride physique-biologique. Cette fonction de coût, grâce à l'expression de ses dérivées premières, a été couplée à une optimisation de la fluence par la méthode BFGS. L'évaluation de cette approche a été réalisée sur les données de dix patients traités pour un cancer de la prostate. Une comparaison à la méthode classique reposant uniquement sur des critères dose-volume a été menée. Elle a montré que la méthode proposée produit des planifications cliniquement satisfaisantes avec de meilleurs résultats en termes de toxicité prédite. L'approche proposée a aussi été comparée à deux autres

méthodes d'optimisation (recuit simulé et optimisation non-linéaire). Cette étude a montré que notre approche fournit de meilleurs résultats en un temps de calcul inférieur. Les perspectives de ce travail consistent notamment en l'extension de l'approche proposée pour optimiser directement les paramètres machine de façon à fournir un plan de traitement réalisable par l'accélérateur.

Lors de la délivrance du traitement, des modifications anatomiques et des erreurs de positionnement du patient peuvent survenir. Elles constituent des incertitudes géométriques qui peuvent être divisées en erreurs aléatoires et erreurs systématiques. Pour compenser les erreurs aléatoires lors de la planification, différentes approches ont été proposées, comme la méthode de simulation stochastique ou la méthode de convolution de la dose. Ces méthodes entraînent cependant une expansion du volume cible clinique (CTV, *Clinical Target Volume*). Pour éviter cette expansion, l'approche du « robust beam profile » a été proposée, reposant essentiellement sur une augmentation de la fluence au bord des faisceaux. Différentes méthodes de déconvolution ont été utilisées pour obtenir le nouveau profil de faisceau, avec cependant des limites, notamment la présence de hautes fréquences autour des bords du faisceau. Dans ce travail, une approche de déconvolution combinant une décomposition en séries de Taylor et un filtrage avec un filtre de Butterworth a été proposée pour compenser les erreurs aléatoires. Elle a été comparée avec deux autres méthodes de déconvolution sur deux cartes de fluence 2D et un cas prostatique. Les résultats ont montré l'efficacité de la méthode proposée, qui permet de réduire les oscillations de haute fréquence ainsi que la présence de points chauds et froids des profils de fluence.

Toujours lors de la délivrance du traitement, en cas de variation anatomique importante, telle une fonte tumorale en radiothérapie ORL, une ou plusieurs replanifications du traitement peuvent s'avérer nécessaires, notamment pour limiter l'irradiation des glandes parotides. En s'appuyant sur des données recueillies lors d'un protocole clinique en ORL (projet ARTIX) impliquant une replanification hebdomadaire, 31 scénarii différents de replanification, suivant le nombre et les moments des replanifications, ont été simulés. Ainsi, pour chaque scénario, les doses reçues par les patients ont été calculées pour chaque semaine suivant la planification considérée dans le scénario (planification initiale ou replanification). Les critères de comparaison ont été extraits de la dose cumulée reçue au cours de l'ensemble du traitement, calculée grâce au recalage élastique entre délinéations des organes à la planification et lors du traitement. Les critères de couverture de la cible tumorale étant systématiquement respectés, l'analyse s'est focalisée sur la dose reçue par les parotides. Les résultats obtenus sur onze patients ont permis d'identifier les scénarii optimaux de replanifications avec notamment un scénario, optimal pour les deux parotides, à trois replanifications réalisées aux semaines 1, 2 et 5. Ce scénario de replanification permet de diminuer significativement la dose reçue par les parotides puisque le risque de toxicité, établi par le modèle NTCP, est réduit de près de 9%. La suite de ce travail porte sur l'étude des corrélations entre descripteurs anatomiques et gain de la replanification de

façon à proposer des critères de déclenchement d'une replanification dans un contexte de radiothérapie adaptative.

En conclusion, nous présentons dans ces travaux différentes méthodes pour l'optimisation des plans de traitement en radiothérapie des cancers de la prostate et ORL. Ces méthodes ont été évaluées et, chacune à son niveau, permettent d'améliorer les plans de traitement. Les perspectives de ce travail concernent la combinaison de ces méthodes dans un processus complet de planification pour évaluer leur impact dans un contexte clinique.

Publications :

Journaux :

[1] Pengcheng Zhang, Jian Zhu, Yuanjin Li, Huazhong Shu. Improved algorithm of multileaf collimator field segmentation. Journal of southeast university (Natural Science Edition), 2012, 42(5):875-879. (en Chinois)

[2] Pengcheng Zhang, Renaud De Crevoisier, Antoine Simon, Pascal Haigron, Jean-Louis Coatrieux, Baosheng Li and Huazhong Shu. A new deconvolution approach to robust fluence for intensity modulation under geometrical uncertainty. Phys Med Biol, 2013, 58(17):6095-6110.

[3] Pengcheng Zhang, Antoine Simon, Renaud De Crevoisier, Pascal Haigron, Mohamed Hatem Nassef, Baosheng Li and Huazhong Shu. A new pencil beam model for photon dose calculations in heterogeneous media, Physica Medica, accepted.

[4] Pengcheng Zhang, Antoine Simon, Huazhong Shu, Pascal Haigron, Caroline Lafond, Jean-Louis Coatrieux, Jian Zhu, Baosheng Li and Renaud De Crevoisier. IMRT optimization including the equivalent convex NTCP constraint, submitted.

Conférence :

[5] Pengcheng Zhang, Huazhong Shu, Antoine Simon, Pascal Haigron and Renaud De Crevoisier. A new pencil beam model for photon dose calculations. The 2013 6th International Conference on BioMedical Engineering and Informatics (BMEI 2013)

Résumé étendu en Anglais

Planning Optimization in Radiation Therapy of Prostate and Head-and-Neck

Chapter I. Introduction.....	6
I.1. Radiotherapy in cancer treatment.....	6
I.2. Considered challenges in radiation therapy plan optimization.....	7
Chapter II. A new pencil beam model for photon dose calculations	10
II.1. Introduction	10
II.2. Kernel computation in a modified spherical coordinate system	11
II.2.1. Modified spherical coordinate system.....	11
II.2.2. Getting the pencil beam kernels.....	12
II.2.3. Extracting the parameters of pencil beam kernels	12
II.3. Dose calculations with superposition method.....	13
II.3.1. Dose calculations in water phantom	13
II.3.2. Dose calculations in heterogeneous media	13
II.4. Dose calculation acceleration using FFTC	16
II.5. Validations of dose calculation methods: comparison with AAA and MC methods	16
II.6. Results	17
II.6.1. Water phantom.....	17
II.6.2. Lung slab phantom.....	18
II.6.3. Bone slab phantom.....	19
II.6.4. Lung and bone block phantoms	20
II.6.5. Computation time.....	21
II.7. Discussion.....	21
II.8. Conclusion	22
Chapter III. IMRT optimization incorporating equivalent convex NTCP constraint	23
III.1. Introduction	23
III.2. Advantages of biological cost functions	24
III.2.1. Limitations of dose-volume-based treatment planning.....	24
III.2.2. Advantages of biological cost functions over dose-volume cost functions.....	25
III.3. Methods and Materials.....	25
III.3.1. Dose calculations.....	25
III.3.2. Equivalent convex NTCP constraint.....	25
III.3.3. Hybrid physical-biological model	26
III.3.4. Optimization method of the fluence elements.....	27
III.3.5. NTCP-based optimizations proposed by Mohan <i>et al</i> and Stavrev <i>et al</i>	28
III.3.6. Organ delineation and dose prescription for 10 patients	29
III.3.7. Software for optimization and comparison test	29
III.4. Results.....	30
III.4.1. Comparison of the proposed method with dose-volume-based optimization	30
III.4.2. Comparison of the proposed method with SA and NP	33
III.5. Discussion	34
III.6. Conclusion.....	36
Chapter IV. The corrections of random geometric uncertainties in IMRT	38

IV.1. Introduction	38
IV.2. Methods and Materials.....	39
IV.2.1. The beam profile deconvolution with the series expansion method.....	39
IV.2.2. 1D beam profile deconvolution	39
IV.2.3. Extension to 2D fluence deconvolution	42
IV.3. Results.....	43
IV.3.1. Results for a regular 2D fluence map	43
IV.3.2. Results for an IMRT field	46
IV.3.3. Results for a prostate case.....	48
IV.4. Discussion and conclusion	51
Chapter V. Assessment of the optimal time and number of replannings in head-and-neck cancer adaptive radiotherapy to spare the parotid glands.....	53
V.1. Introduction	53
V.2. Purpose and general framework of the study	53
V.3. Materials.....	55
V.3.1. Patients and tumors.....	55
V.3.2. Treatment and planning.....	55
V.3.3. Weekly dose estimations in cases of replanning	56
V.4. Methods for dose accumulation.....	56
V.4.1. Registration algorithm.....	56
V.4.2. Aligning the dose distributions.....	57
V.4.3. Calculating the accumulated dose.....	58
V.4.4. Analyzing the cumulated dose and estimating the optimal scenario	58
V.5. Results.....	59
V.5.1. Anatomical variations occurring during the 7 weeks of treatment	59
V.5.2. Accumulated parotid gland dose distributions without replanning (compared to the planning)	60
V.5.3. Accumulated parotid gland dose distributions corresponding to 31 scenarios of replannings	61
V.5.4. Optimal number and time of replannings.....	64
V.5. Discussion	65
V.6. Conclusion.....	66
Chapter VI. Conclusions	67
References	69

Abbreviations

AAA	Anisotropic analytical algorithm
ART	Adaptive radiotherapy
CT	Computed tomography
CTV	Clinical target volume
DV	Dose-volume
DVH	Dose volume histogram
EUD	Equivalent uniform dose
FFTC	Fast Fourier transform convolution
H&N	Head-and-Neck
ICRU	International Commission on Radiation Units and Measurements
IMRT	Intensity-modulated radiation therapy
L-BFGS	limited-memory Broyden-Fletcher-Goldfarb-Shanno algorithm
MC	Monte Carlo
NTCP	Normal tissue complication probability
PG	Parotid gland
PTV	Planning target volume
RT	Radiation therapy
SRS	Stereotactic radiosurgery
TCP	Tumor control probability
TPS	Treatment planning system

Chapter I. Introduction

In this introduction, we briefly present the crucial role of radiotherapy in the treatment of cancer. We introduced afterwards the addressed issues in the different aspects of radiation therapy that we considered in this work: pencil beam dose calculations, intensity modulated radiation therapy (IMRT) optimizations, geometric uncertainties in IMRT, and adaptive radiotherapy (ART) of head-and-neck cancers.

I.1. Radiotherapy in cancer treatment

Cancer has become a common disease which seriously threatens the human health. According to recent estimates of the International Agency for Research on Cancer (IARC) and World Health Organization (WHO), approximately ten million new cancer cases are being detected per year world-wide, with slightly more than half of the cases occurring in developing countries ^[1]. By the year 2015 this number is expected to increase to about 15 million cases, among which two thirds will occur in developing countries. The cancer prevention in China is a challenging task. The cancer mortality rate has increased by nearly 30% in the past 20 years ^[2]. In 2008, new cancer cases reached 2.82 million, and this number will increase to 21 million in 2030.

Radiation therapy is one of the three major means for tumor treatment, together with surgery and chemotherapy. According to recent estimates of the WHO, the tumor cure rate is about 55% (with a high variability depending on the tumor), whereas the contribution of radiation therapy is up to 40% to cure the patients ^[2]. Compared to surgery, radiation therapy conserves the integrity of the organ affected by the tumor and aims also to keep the organ functionality to limit the side effects and maintain a high quality of life after the treatment.

The radiations used for cancer treatment are ionizing radiation, which can be grouped into two major types: photons (X-rays and gamma rays), which are most widely used in cancer treatment, and particle radiation (such as electrons, protons, neutrons, carbon ions, alpha particles, and beta particles). In this thesis, we mainly focus on the radiotherapy of external photon beams. Ionizing radiation damages the DNA (DNA double strand breaks) of the cells within both normal tissues and tumors, and makes them unable to divide and grow ^[3]. Although radiation is directed at the tumor, the surrounding normal tissues are also inevitably affected by the radiation and therefore damaged. However, the normal tissues and tumors are prone to different biological reactions to the radiation. Radiation therapy is based on the differential effect of ionizing radiations, killing the tumor cells and preserving in a certain extent the healthy tissue due to the ability of the normal cells to repair the DNA double stand breaks. The goal of radiation therapy is therefore to maximize the dose delivered to tumors while minimizing exposure to healthy tissues, i.e. improving tumor control probability (TCP) while reducing the normal tissue complication probability (NTCP).

A radiotherapy treatment can be divided in two main steps: (i) the treatment planning during which the treatment parameters are defined and optimized according to the prescription and the patient's anatomy; (ii) the treatment delivery, fractionated in a series of daily fractions.

Advances in imaging technologies (CT, MRI and PET), dose calculation algorithms (3D superposition, Monte Carlo) and dose delivery techniques (IMRT, SRS, IGRT) have made it possible to deliver the prescribed dose in a high conformal fashion to the treatment volume, while sparing relatively well surrounding normal tissue. Radiation therapy has therefore gradually been developed towards a “precise identification and characterization of the tumor, precise planning, and precise treatment”^[4]. The treatment plan optimization is performed by the treatment planning system (TPS) which needs a lot of key technologies, such as rapid 3D reconstructions, high-precision 3D dose calculations, plan optimizations and inverse planning.

This thesis focused on different challenges concerning plan optimization methods.

I.2. Considered challenges in radiation therapy plan optimization

Many methods have been proposed to improve the quality of IMRT plans considering different aspects, such as dose calculations, plan optimizations, geometric uncertainty corrections and ART techniques. However, several challenges are remaining, including the following ones (cf. Fig. 1.1).

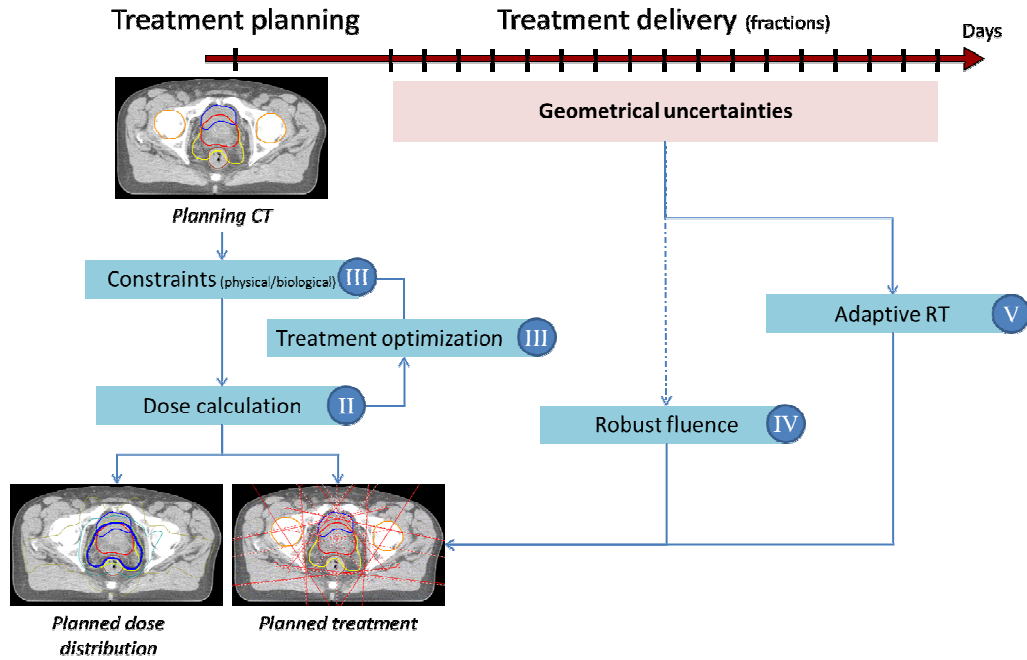


Figure 1.1 Global workflow of RT treatment and position of the different parts of the work. The surrounded numbers are the corresponding chapters.

Dose calculation algorithms play a crucial role in modern TPS. It creates the link between the chosen treatment parameters, and the observed clinical outcome for a specified treatment technique. The ICRU recommends that the dose to the radiation therapy target volume should be delivered with an accuracy of 5% or better ^[5]. An improvement of 1% in dose accuracy may result in an increase of 2% in cure rate for early stage tumors ^[6]. Hence, some sophisticated methods have been developed to improve the accuracy of dose calculations, such as kernel based methods and Monte Carlo (MC) methods. However, the higher is the accuracy of the algorithm, the more computation time is needed for dose calculations. In inverse planning optimization, dose calculations might be repeated many times (usually from 10's to 1000's) ^[7]. The pencil beam approach is frequently used in dose calculations to achieve a good compromise between speed and accuracy, which is one of the crucial challenges for the development of modern dose calculation algorithms. However, this pencil beam approach suffers from limitations. Firstly, if the rays from the source are divergent, most pencil beam models, consider parallel beamlets, resulting to imprecision. Secondly, the correction of heterogeneities involves approximations, especially considering the depth-directed components. In this part of the work, we proposed to improve the pencil beam algorithm, without increasing the computation time, by considering a modified spherical coordinate system and by modifying the heterogeneities correction method. Moreover, in order to accelerate the computation, we considered the use of the Fast Fourier Transform to perform the convolutions.

Still in the planning step, the most commonly used approach in IMRT planning is the optimization approach, in which the best physically and technically possible treatment plan is iteratively searched. Optimization is mathematically defined as the maximization or minimization of a score incorporating different criteria. These criteria can either be formulated as physical criteria or as biological criteria ^[8]. Physical criteria are often implemented as a dose-dependent function calculating the mean-squared deviation from a prescribed dose. However, physical criteria are more often unable to accurately render the nonlinear response of tumors or normal structures to irradiation, especially with arbitrary inhomogeneous dose distributions. Compared to physical criteria, biological criteria modeling the biological response to a given dose are potentially more directly associated with treatment outcome. However, biological criteria are often sigmoidal functions of dose and hence, are inherently non-linear and non-convex ^[9]. Thus, the direct incorporation of such non-linear and non-convex criteria into objective function makes the global optimum extremely difficult to find. We then considered an equivalent convex NTCP constraint, and its incorporation in IMRT optimization.

During the treatment delivery, some geometrical uncertainties hamper the precision of the delivery of the plan. The ICRU considers three sources of geometrical uncertainty: patient set-up variation, organ motion and deformation, and machine related errors ^[10]. A lot of effort has been put in order to measure and reduce these geometrical uncertainties which can be classified as random and systematic ^[11, 12]. Systematic uncertainties can be reduced by image-guidance techniques and adaptive radiotherapy ^[12-15]. Large organ movements ^[16], such as those observed for instance in lung or in prostate cancers, can be compensated by real-time motion tracking. However, random uncertainties cannot be

fully eliminated ^[17], due to the finite response time and irregular motion patterns ^[18]. Thus, in order to reduce random uncertainties, we proposed to use an approach based on the deconvolution of the beam profile using a series expansion and Butterworth filtering.

During the treatment delivery, the location, shape, and size of target and normal tissue may be altered significantly due, for example, to weight loss and organs' responses to the radiation during a 6-7 weeks course of treatment ^[19-22]. This is especially the case in the treatment of head-and-neck (H&N) cancers. During the treatment, these changes could lead to unexpected high normal tissue complications and low tumor control, even by using the initial highly conformal IMRT plan ^[19, 23, 24]. Therefore, the plan created on the initial planning Computed Tomography (CT) may no longer be optimal. Adaptive radiation therapy (ART) techniques may be therefore used to adjust the initial plan and to effectively correct these variations by replanning(s). Replanning is a heavy task, which includes acquiring CT scans, delineating organs, optimizing a new plan, up to the associated quality control before the treatment. Due to the limited resources in the hospitals, it appears almost impossible to systematically replan a patient every week. However, within a clinically acceptable range, the number of replannings should be reduced properly in function of the tumor. Therefore, it is necessary to determine the optimal number and time of replannings in an ART workflow. In this work, we proposed to simulate different strategies in H&N ART, in order to select the optimal strategy.

The following four chapters describe all these challenges, the proposed methods and the obtained results. Based on the existing studies on dose calculation, the first chapter describes how the standard pencil beam algorithm was improved to increase the accuracy of dose calculations. It corresponds to a paper which has been accepted for publication in *Physica Medica*. In the second chapter, we describe the IMRT optimization including the equivalent convex NTCP constraint, corresponding to a paper submitted to *Physica Medica*. Then, in the third chapter, we propose the deconvolution approach to robust fluence for intensity modulation under geometrical uncertainty. This work has been published in *Physics in Medicine and Biology* in 2013. In the last chapter, we describe the protocol and the results we considered to simulate different ART workflow in H&N radiation therapy. In the last part we conclude this work before drawing some perspectives.

This research work has been performed within different laboratories:

- Laboratory of Image Science and Technology, Southeast University, Nanjing 210096, People's Republic of China
- Laboratoire Traitement du Signal et de l'Image, Université de Rennes I, INSERM, U1099, Rennes F-35042, France
- Centre de Recherche en Information Médicale Sino-français (CRIBs), Rennes F-35042, France
- Shandong Tumor Hospital, Jinan 250117, People's Republic of China
- Centre régional de lutte contre le cancer Eugène Marquis, Rennes F-35000, France

Chapter II. A new pencil beam model for photon dose calculations

II.1. Introduction

A fast and accurate calculation of 3D dose distributions is one important issue of modern radiation oncology. It is essential, particularly for intensity-modulated radiation therapy (IMRT), where physicists and physicians rely heavily on dose calculation accuracy, in order to optimize and evaluate therapy planning. Inverse planning optimization makes further demands of dose calculations, as it is an iterative process in which dose calculations might be repeated tens to thousands of times^[7]. The pencil beam approach is commonly used for dose calculations in order to achieve an optimum compromise between speed and accuracy, which is one of the crucial challenges for developing modern dose calculation algorithms.

The concept of pencil beam algorithms using pencil kernels can be described as the convolution/superposition of some unit photon fluence of a mono-directional beam with the pencil beam kernel modeled in water. The potential of pencil beam algorithms for dose calculations in radiotherapy has been exploited and investigated by various research groups^[7, 25-34]. Some of them have investigated the use of fast transform convolution techniques in order to accelerate dose calculation, such as the fast Fourier transform convolution (FFTC)^[35] and the fast Hartley transform^[36]. Nevertheless, different approximations have been made for pencil beam algorithms, for the most part related to primary beam spectral variations, beam divergence, and tissue heterogeneity density scaling^[37].

Firstly, kernel tilting is neglected in most pencil beam algorithms. The rays from the source are divergent but, in most pencil beam models, are considered as parallel beamlets, with the 3D dose distributions directly calculated by parallel kernels^[38]. Transforming the divergent pencil beam kernel for each beamlet in Cartesian coordinates is, however, time-consuming. Some authors have proposed using spherical coordinate systems, such as a diverging coordinate system^[25, 26] and beam's eye view coordinate system^[27, 28]. This would enable to avoid rotating each kernel and save on computation time. In these two coordinate systems, the voxels become larger, however, as they move away from the source, thus affecting dose calculation accuracy. In this part of work, we have modified the coordinate system in order for all the voxels to be of the same size thus limiting the error induced by different voxel sizes.

Secondly, the correction of heterogeneities involves approximations. For the anisotropic analytical algorithm (AAA)^[26], the depth-directed components are usually corrected according to the entry points' positions on the surface. The densities in the previous layers are partially considered, while the lateral components are corrected by a scale factor containing depth-directed information^[25, 26, 39]. This reduces the dose calculations' accuracy. In order to overcome this drawback, we chose to correct the depth-directed components according to the interaction point, where the beamlet interacts

with the given spherical shell. The lateral components were thus calculated directly, without taking the densities from the previous layers into account.

Thirdly and finally, dose calculations are time-consuming. Significant computation time can be saved by performing FFTC calculations on each spherical shell. For the AAA algorithm^[26], the depth-directed components are corrected according to the beamlet entry point on the surface. None of the calculation points on a shell can then be located on the same shell following depth direction correction. Moreover, the depth information in lateral components cannot be separated from lateral information, which hinders getting the invariant convolution kernel for the application of FFTC. In terms of correcting depth-directed components, the AAA algorithm^[26] cannot therefore be directly accelerated with FFTC. To solve this shortcoming, we decided to correct the depth-directed components according to the interaction point. In the FFTC calculations, we used the invariant kernel modeled in the water phantom directly on each spherical shell. Still, with this method the density changes in the lateral direction cannot be taken into account, which results in reduced dose calculation accuracy.

In this part of the work, the dose distribution was calculated in the spherical coordinate system where all the voxels exhibited the same arc length, in order to address kernel tilting. For the issue of heterogeneities, the depth-directed components were corrected according to the interaction point. The lateral components were calculated directly, without taking into account the density changes from the previous layers. In terms of correcting depth-directed components, we were thus able to accelerate the dose calculations directly with FFTC on each spherical shell. The proposed methods have been evaluated on different phantom geometries, representing realistic situations. The dose distributions calculated by the proposed method have been then compared to those obtained with the Monte Carlo (MC) method and the AAA algorithm.

II.2. Kernel computation in a modified spherical coordinate system

II.2.1. Modified spherical coordinate system

In classic dose calculations, an orthonormal basis is used for the kernel coordinate system. Its origin is located at the radiation source, with the positive z-axis passing through the isocenter. The x- and y-axes are aligned with the corresponding collimator axes. The sampling method with fixed zenith angle θ_0 is that which is typically used in the spherical coordinate system, such as the AAA algorithm. A spherical mapping from the kernel coordinate system to spherical coordinate system is then defined as follows:

$$\mathbf{x} \mapsto \left(\frac{\arctan(\mathbf{x}_x / \mathbf{x}_z)}{\theta_0}, \frac{\arctan(\mathbf{x}_y / \mathbf{x}_z)}{\theta_0}, \sqrt{\mathbf{x}_x^2 + \mathbf{x}_y^2 + \mathbf{x}_z^2} \right) := \mathbf{p}, \quad (2.1)$$

where $\mathbf{x} = (\mathbf{x}_x, \mathbf{x}_y, \mathbf{x}_z)$ is a point in the Cartesian coordinate system, $\mathbf{p} = (\mathbf{p}_x, \mathbf{p}_y, \mathbf{p}_z)$ is the corresponding point in the spherical coordinate system, and θ_0 is the sampling angle of zenith angle. For a fixed zenith value θ_0 , however, the voxel gradually becomes larger as it moves away from the origin, which results in a decrease in dose calculation accuracy with

the increasing voxel sizes. In order to avoid errors caused by different voxel sizes, we have proposed ensuring all the voxels exhibit the same arc length. The new spherical mapping $M: \mathcal{R}_+^3 \rightarrow \mathcal{R}_+^3$ from the kernel coordinate system to modified coordinate system is thus given as:

$$\mathbf{x} \mapsto \left(\frac{\arctan(\mathbf{x}_x / \mathbf{x}_z)}{\Delta} \mathbf{p}_z, \frac{\arctan(\mathbf{x}_y / \mathbf{x}_z)}{\Delta} \mathbf{p}_z, \mathbf{p}_z \right) := \mathbf{p}, \quad (2.2)$$

where Δ is the fixed value for arc length and $\mathbf{p}_z = \sqrt{\mathbf{x}_x^2 + \mathbf{x}_y^2 + \mathbf{x}_z^2}$.

II.2.2. Getting the pencil beam kernels

Energy deposition kernels can be accurately calculated by means of the MC method [40]. In our study, the heterogeneities were corrected according to the beamlet interaction point in the phantom, as well as by changing the position of the entry point for the given beamlet, inducing variation in the distances from source to surface (DSS) for this beamlet. In the modified spherical coordinate system, the pencil beam kernels are not the same for different DSSs, and each corresponding kernel therefore needs to be found. To achieve this, the pencil beam kernels must be re-sampled into the spherical coordinate system for the different DSSs.

In order to reduce dose computation time, the pencil beam kernel was expressed as a poly-energetic kernel, which is a superposition of mono-energetic kernels weighted by the spectrum of the beamlet. The poly-energetic kernel was mapped directly into the modified spherical coordinate system using the mapping M defined in Eq. (2.2). The pencil beam kernels of different DSS values were then defined in the modified spherical coordinate system as:

$$h_{DSS}(\mathbf{p}) = h_{cyl}(\mathbf{x}) \det(J(M(\mathbf{x}))), \quad (2.3)$$

where h_{cyl} is poly-energetic kernel, $\mathbf{x} = M^{-1}(\mathbf{p})$ is the considered point in the Cartesian system, J is the Jacobian matrix of the mapping M in the modified spherical coordinate system, and $\det(J(M(\mathbf{x})))$ is the determinant of the Jacobian matrix J , which is used to account for the non-uniform volume mapping of the M operator as follows:

$$\det(J(M(\mathbf{x}))) = \frac{\mathbf{x}_z (\mathbf{x}_x^2 + \mathbf{x}_y^2 + \mathbf{x}_z^2)^{3/2}}{(\mathbf{x}_x^2 + \mathbf{x}_z^2)(\mathbf{x}_y^2 + \mathbf{x}_z^2)}. \quad (2.4)$$

II.2.3. Extracting the parameters of pencil beam kernels

The pencil beam kernels were modeled as a function of photon energy, with respect to a depth measurement in the depth direction and radius measurement in the lateral direction. The function's parameters were extracted from the MC-derived data and stored in a database. In standard practice, the pencil beam kernel derived from MC simulations is separated into depth-directed and lateral components. In our study, the depth-directed component accounted for the total energy deposited by the pencil beam for a spherical shell at depth \mathbf{p}_z , as seen below:

$$I_{DSS}(\mathbf{p}_z) = \Phi \iint h_{DSS}(\mathbf{p}_x, \mathbf{p}_y, \mathbf{p}_z) d\mathbf{p}_x d\mathbf{p}_y, \quad (2.5)$$

where Φ is the primary energy fluence at the phantom surface and h_{DSS} is the kernel of a given DSS. The lateral components accurately modeled the lateral scatter on the spherical shell. For the spherical shell at depth \mathbf{p}_z , the fraction of energy $f_{DSS}(r, \mathbf{p}_z)$ deposited onto an infinitesimally small angular sector at radius r from the beamlet central axis was calculated based on the MC-derived data:

$$f_{DSS}(r, \mathbf{p}_z) = h_{DSS}(r, \mathbf{p}_z) / I_{DSS}(\mathbf{p}_z). \quad (2.6)$$

In order to accelerate dose calculation with FFTC (in Sec. 2.3), the depth information had to be easily separated from lateral information for lateral components to get the invariant convolution kernel for the application of FFTC. We therefore used the function proposed by Tillikainen *et al.* ^[26] in order to model the lateral scatter (Eq. (2.6)). The lateral pencil-beam component was defined as the sum of multiplications of the depth $c_{DSS,i}(\mathbf{p}_z)$ and lateral information, as shown below:

$$k_{DSS}(r, \mathbf{p}_z) = \sum_{i=1}^{N=6} c_{DSS,i}(\mathbf{p}_z) \frac{1}{\mu_i} e^{-\mu_i r}, \quad (2.7)$$

where the attenuation coefficients μ_i are the same for all spherical shells to ensure efficient implementation. The fitting algorithm was used to extract parameters $c_{DSS,i}(\mathbf{p}_z)$ in Eq. (2.7). We stored depth components $I_{DSS}(\mathbf{p}_z)$, attenuation coefficients μ_i , and weight parameters $c_{DSS,i}(\mathbf{p}_z)$ in a database for each kernel of different DSS values.

II.3. Dose calculations with superposition method

II.3.1. Dose calculations in water phantom

It is well-known that particles follow different paths to pass through media. For the pencil beam methods, particles are first assumed to arrive at the destination spherical shell p_z via the central line of a beamlet, then traveling to the destination voxels along a few collapsed paths on the spherical shell ^[26]. In a homogeneous water-equivalent phantom, the energy $E(\mathbf{p})$ deposited by a pencil-beam beamlet into grid point \mathbf{p} is calculated by multiplying the energy $I_{DSS}(\mathbf{p}_z)$ deposited on the calculation spherical shell at depth \mathbf{p}_z with the corresponding lateral scatter kernel k_{DSS} , as follows:

$$E(\mathbf{p}) = I_{DSS}(\mathbf{p}_z) k_{DSS}(r, \mathbf{p}_z), \quad (2.8)$$

where $I_{DSS}(\mathbf{p}_z)$ is the depth-directed component defined in Eq. (2.5), k_{DSS} is the total lateral component defined in Eq. (2.7), and r is the distance from point p to the interaction point on a given spherical shell. The equivalent depth d was calculated by ray tracing the beamlet's central axis through the phantom from the interaction point to phantom surface. The corresponding DSS is defined as $\mathbf{p}_z - d$. Other parameters in Eq. (2.8) could easily be taken from the previously-compiled database.

II.3.2. Dose calculations in heterogeneous media

In heterogeneous media, tissue heterogeneities are corrected by the equivalent path length (EPL), which scales the path length according to the relative electron densities

between tissue and water ^[41]. In a patient, the EPL of the actual path was defined by the following spatially-dependent scaling factor:

$$d_{eff}(X) = \int_X \rho^{elec}(p) / \rho_{water}^{elec} dp, \quad (2.9)$$

where X is the actual path, ρ^{elec} is the local electron density at point p , and ρ_{water}^{elec} is the electron density of water.

The pencil beam kernels were separated out into depth-directed and lateral components. We corrected the heterogeneity effects by scaling the depth-directed and lateral components. In typical practice, depth-directed components are corrected by moving both the interaction and calculation points (Fig. 2.1(a)). This can, however, induce some error in the spherical coordinate system. As displayed in Fig. 2.1(a), calculation point A must, in fact, be moved to point A'' , not point A' , when interaction point B is moved to point B' , due to the arc length and curvature of AB not being equal to the arc length and curvature of $A'B'$. For the AAA algorithm ^[26], point A'' is replaced, in an imprecise manner, with point A' by multiplying arc length AB with a scaling factor. In order to avoid this lack of precision, we proposed moving the beamlet entry point on the surface and maintaining the point A and B positions fixed, as displayed in Fig. 2.1(b).

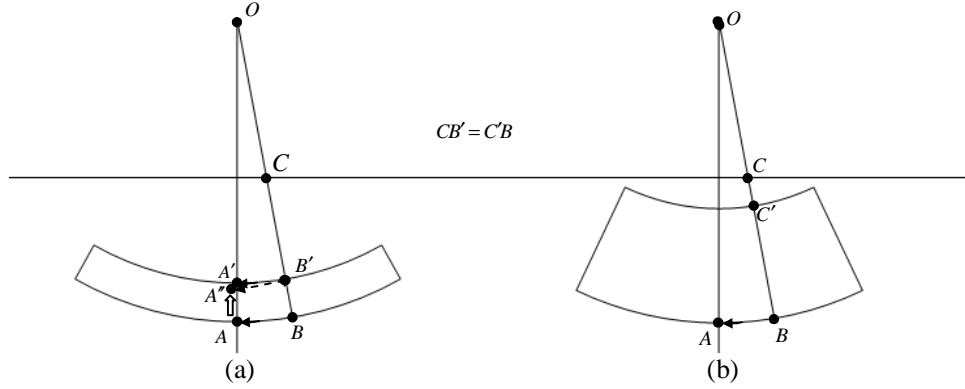


Figure 2.1 Two different manners for correcting tissue heterogeneities: (a) changing the position of interaction point B , (b) moving the entry point C to point C' and changing the SSD of the beamlet.

For the depth-directed component, therefore, function I_{DSS} was corrected by moving the beamlet entry point on the surface, thus achieving the DSS variation for the given beamlet. The new DSS value was computed as follows:

$$SSD' = \mathbf{p}_z - deff(P), \quad (2.10)$$

where P is the path from the beamlet entry point to interaction point \mathbf{p}' . The depth-directed component was corrected by using the DSS' kernel. The depth-directed component also required scaling by the local electron density, with the correction for the depth-directed component thus given as:

$$I'_{DSS'}(\mathbf{p}_z) = I_{DSS'}(\mathbf{p}_z) \rho_w(\mathbf{p}'), \quad (2.11)$$

where $I_{DSS'}$ is the depth-directed component for the DSS' kernel and $\rho_w(\mathbf{p}')$ is the relative electron density at interaction point \mathbf{p}' .

The computation of lateral components was conducted in a similar fashion. The

energy released at interaction point \mathbf{p}' was transported and deposited on the spherical shell at depth \mathbf{p}_z . The effective radius r' was calculated by following collapsed path $C(\mathbf{p})$ from interaction point \mathbf{p}' to calculation point \mathbf{p} on the given spherical shell. The heterogeneity-corrected lateral component $k'_{DSS'}(r, \mathbf{p}_z)$ was thus expressed as follows:

$$k'_{DSS'}(r, \mathbf{p}_z) = k_{DSS'}(r', \mathbf{p}_z) \rho_w(\mathbf{p}), \quad (2.12)$$

where $k_{DSS'}$ is the lateral component for the DSS' kernel and r' is the effective radius computed as $r' = \text{deff}(C(p_z))$. Compared with the AAA algorithm, our proposed method does not require the effective radius to be multiplied with a scale factor owing to the depth-directed correction method we used.

The heterogeneity-corrected energy distribution from a single beamlet was calculated as the multiplication of a depth-directed component (Eq. (2.11)) with a lateral component (Eq. (2.12)):

$$E(\mathbf{p}) = I'_{DSS'}(\mathbf{p}_z) k'_{DSS'}(r, \mathbf{p}_z). \quad (2.13)$$

The total energy deposited into a grid point \mathbf{p} is the integral of the contributions from the individual beamlets over the broad beam area:

$$E_{tot}(\mathbf{p}) = \iint_S E(\mathbf{p}) dS, \quad (2.14)$$

where S is the set of all beamlets. The “collapsed cone” method ^[42] was applied for the superposition of the contributions from all individual beamlets. In our experiments, it was assumed that the energy released at the interaction point was transported into eight equal angular sectors, over which the superposition was performed. With respect to the x-axis of the spherical coordinate system, the cone axis directions were 0° , 45° , 90° , 135° , 180° , 225° , 270° , and 315° for the eight angular sectors, respectively.

Given that the absorbed dose is defined as the imparted energy per unit mass, the dose to a point at \mathbf{x} in the Cartesian coordinate system is given by:

$$D(\mathbf{x}) = \frac{E_{tot}(\mathbf{p})}{\rho_w(\mathbf{x}) \det(J(M(\mathbf{x})))}, \quad (2.15)$$

where J is the Jacobian of M and $\mathbf{p} = M(\mathbf{x})$ is the corresponding position in the modified spherical coordinates for the point at \mathbf{x} . In order to decrease the error caused by the coordinate transform between these two coordinate systems, we applied the linear interpolation algorithm for the purposes of calculating the energy deposited at a point \mathbf{p} .

In this study, we commenced by obtaining the pencil beam kernel by means of an MC simulation in the Cartesian coordinate system. This kernel was then re-sampled into the modified spherical coordinate system by applying the different sampling methods. These sampling methods differed in terms of the assumed source positions, resulting in different DSSs, and corresponded to the kernels of different DSSs. Each sampling method can be used in order to estimate the dose deposited at a certain point from the given beamlet. In this way, heterogeneities can be corrected by selecting a kernel of a special DSS, thus avoiding changes in interaction and calculation point positions (Fig. 2.1). The pencil beam kernels are not the same for different DSSs in the modified spherical coordinate system. We modeled kernels of DSS values ranging from 90cm to 115cm. The sampling interval of DSS value was 0.25cm. After obtaining the DSS value, a nearest-neighbor interpolation algorithm was applied to determine the corresponding kernel.

II.4. Dose calculation acceleration using FFTC

In order to calculate dose distributions with fast transform techniques, depth information $I_{DSS}(p)c_{DSS,i}(p)$ was separated from lateral information in Eq. (2.8). This allowed for the total energy deposited on the shell at depth p_z to then be directly obtained by performing the Inverse Fast Fourier Transform (IFFT) of the sum of FFTC, as described below:

$$E_{tot}(\mathbf{p}) = F^{-1} \left\{ \sum_{i=1}^{N=6} F \{ I_{SSD}(\mathbf{p}) c_{SSD,i}(\mathbf{p}) \} F \left\{ \frac{1}{\mu_i} e^{-u_i \sqrt{\mathbf{p}_x^2 + \mathbf{p}_y^2}} \right\} \right\}. \quad (2.16)$$

II.5. Validations of dose calculation methods: comparison with AAA and MC methods

In order to assess our proposed methods, dose distributions were calculated in similar synthetic phantoms to those used by Tillikainen *et al.* ^[26], namely (1) a water phantom, (2) a lung slab phantom consisting of a water phantom with a 15cm slab of lung ($\rho_w=0.3\text{g/cm}^3$), (3) a bone slab phantom consisting of a water phantom with a 5cm slab of bone ($\rho_w=1.85\text{g/cm}^3$), (4) a lung block phantom consisting of a water phantom with a 10cm-thick block of lung positioned 2cm off the central axis (CAX), and (5) a bone block phantom consisting of a water phantom with a 5cm-thick block of bone positioned 2cm off the CAX.

A source-to-phantom distance of 1000mm was used for all field sizes and phantoms. The source model used in the calculations was only the primary source, since our study primarily sought to model the primary beam. We modeled dose distributions for a 6MV photon beam. The 6MV photon beam spectrum was obtained from the EGS4 Spectra Library (file mohan6.spectrum ^[43]). Our proposed pencil beam model (superposition method in Sec. II.2 and FFTC method in Sec. II.3) had been evaluated by comparing depth-dose curves and cross profiles with results obtained from using the AAA and MC methods in the test phantoms for different photon fields. The DOSXYZnrc ^[44, 45] radiation transport code was applied for MC dose calculations. The following parameters were set for the simulation: grid size=0.5cm, ECUT=0.70MeV (electron/positron minimum transport energy), and PCUT=0.01MeV (photon minimum transport energy). One billion particle histories were simulated for each experiment in order to maintain a statistical standard uncertainty of under 0.2% for an overall statistical uncertainty. The AAA algorithm we chose to employ was implemented by modeling the method proposed by Tillikainen *et al.* ^[26]. Both the proposed method and the AAA algorithm calculated the dose distributions in a spherical coordinate system. For our proposed method, all the voxels exhibited the same arc length (in Eq. (2.2)), namely $\Delta = 0.5\text{cm}$, on each spherical shell, whereas the sampling angle was fixed (in Eq. (2.1)) for the AAA method at $\theta_0 = 0.29^\circ$. The spherical shell thickness was 0.5cm for these two methods. Finally, all dose distributions were transformed into a Cartesian coordinate system, with a voxel size of $0.5 \times 0.5 \times 0.5\text{cm}^3$.

II.6. Results

II.6.1. Water phantom

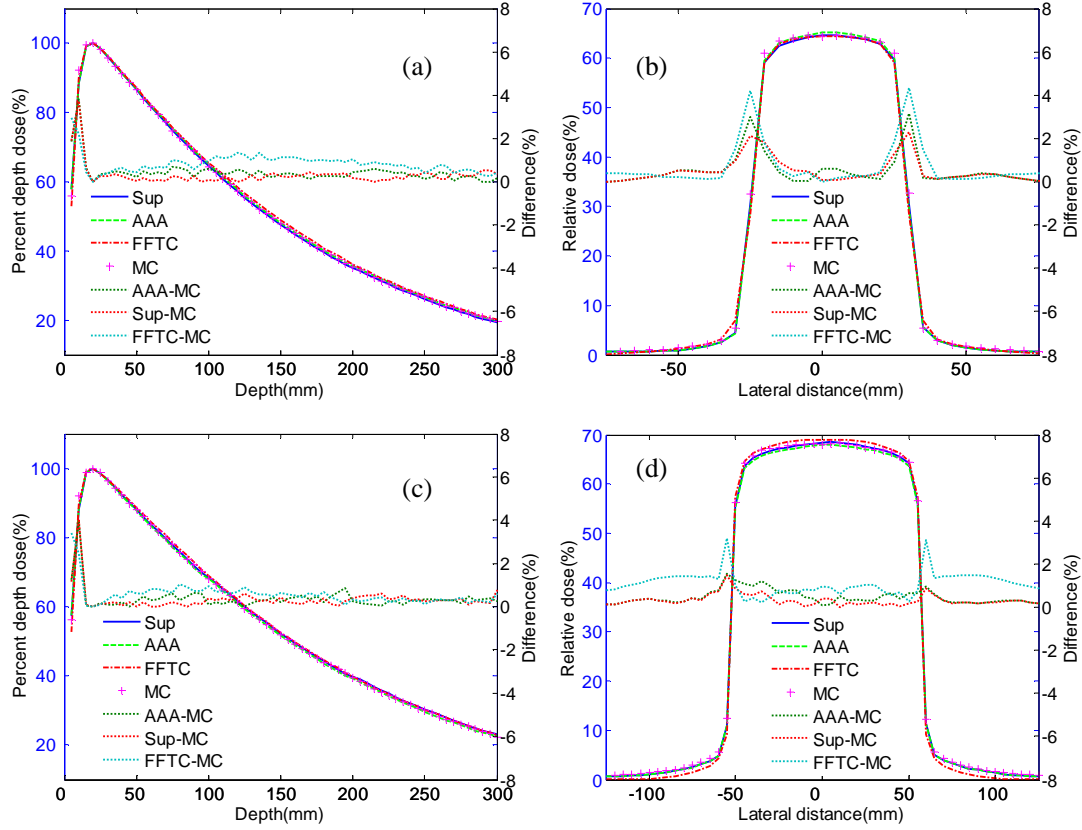


Figure 2.2. Water phantom: comparison of the superposition method (Sup), FFTC method (FFTC), AAA algorithm, and MC simulation. Solid lines represent results calculated using Sup, dashed lines represent results using AAA, dashed and dotted lines represent results using FFTC, and ‘plus’ symbols represent results using MC. Dotted lines correspond to the absolute difference between two calculation methods. (a) Depth-dose curves of $50 \times 50 \text{ mm}^2$ field sizes, (b) profiles at 100mm depth of $50 \times 50 \text{ mm}^2$ field size, (c) depth-dose curves of $100 \times 100 \text{ mm}^2$ field size, and (d) profiles at 100mm depth of $100 \times 100 \text{ mm}^2$ field size.

In the water phantom, we considered depth-dose curves for 50×50 and $100 \times 100 \text{ mm}^2$ field sizes, as well as for corresponding lateral dose profiles at 100mm depth. Figure 2.2 displays the depth-dose curves and lateral dose profiles calculated by means of the superposition method, the FFTC method, the AAA algorithm, and the MC method. The results revealed that the superposition method was more accurate than the AAA algorithm and FFTC method. The mean absolute deviations of the superposition method were 0.31% and 0.34% for depth-dose curve, and 0.41% and 0.33% for lateral dose profiles, respectively for the two fields. The largest differences recorded in the calculated profiles were approximately 2.30% and 1.48%, respectively for the two fields, around the field edges. As we can infer from Fig. 2.2 (b) and (d), the accuracy of lateral dose profiles was

improved by increasing the field sizes for the superposition method. The electronic disequilibrium on the field central axis reduced in size as the field size increased, thus rendering it easier to calculate dose distributions with rectilinear kernel scaling approaches and improving dose calculation accuracy. For the FFTC method, mean absolute deviations were 0.84% and 0.56% for depth-dose curves, and 0.62% and 1.09% for lateral dose profiles, respectively for the two fields. For all phantoms considered in our study, comparisons have been summarized in Table 2.1 for superposition method, FFTC method, and AAA algorithm. Our proposed method could not fully correct the build-up effects near the interfaces where the dose would jump abruptly to a new equilibrium level. Significant maximum deviations in depth-dose curves were still observed in the build-up region, estimated at 3.80% and 3.96% for AAA algorithm, 3.72% and 4.05% for the superposition method, and 2.96% and 3.43% for the FFTC method, respectively for these two fields. We can see in Fig. 2.2 (a) and (c) that, as the field size increased, accuracy decreased for the depth-dose curves in the build-up region. The pencil beam kernel was incapable of satisfactorily taking build-up effects into account. The larger the field size, the more pencil beam kernels were considered for dose calculations, which increased the number of errors occurring in the depth-dose curves in the build-up region.

	Depth curves			Lateral curves		
	Water	Lung slab	Bone slab	Water	Lung block	Bone block
AAA	0.37%	1.59%	0.83%	0.48%	1.72%	0.88%
Superposition	0.34%	0.83%	0.48%	0.33%	0.94%	0.57%
FFTC	0.56%	2.89%	1.05%	1.09%	1.89%	1.08%

Table 2.1. Mean absolute deviations for the AAA algorithm, superposition method, and FFTC method for depth curves in water, and lung and bone slab phantoms, as well as lateral curves in water, and lung block and bone block phantoms, respectively, for 100×100mm² field size.

II.6.2. Lung slab phantom

In the lung slab phantom, we calculated the depth-dose curves for the 50×50 and 100×100mm² field sizes. Depth-dose curve comparisons are presented in Fig. 2.3. The superposition method accurately modeled the attenuations in the lung-equivalent material, as the mean absolute deviations for these two fields were 0.73% and 0.83%, respectively for the two fields. Compared to the AAA algorithm, the superposition method reduced mean absolute deviations by 0.77% and 0.75% for these two fields, respectively. Nevertheless, the interface effects were slightly overestimated, leading to large dose deviations near the heterogeneity boundary of approximately 3.15% and 2.73%, respectively for the two fields. Owing to a neglect of density changes in lateral direction for the FFTC method, significant mean absolute deviations were observed for depth-dose curves, namely 2.85% and 2.89%.

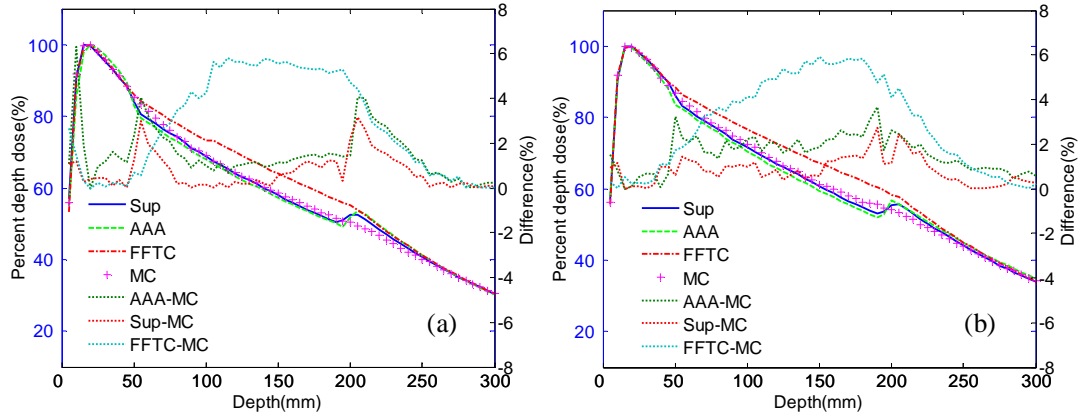


Figure 2.3. Lung slab phantom: comparison of the superposition method (Sup), FFTC method (FFTC), AAA algorithm, and MC simulation. (a) Depth-dose curves of $50 \times 50 \text{mm}^2$ field sizes; (b) depth-dose curves of $100 \times 100 \text{mm}^2$ field size.

II.6.3. Bone slab phantom

Depth-dose curve comparisons pertaining to the bone slab phantom are presented in Fig. 2.4 for the 50×50 and $100 \times 100 \text{mm}^2$ field sizes. For the superposition method, the mean absolute deviations of depth-dose curves were 0.83% and 0.48%, respectively for the two fields, in comparison with 1.01% and 0.83% for the AAA algorithm. We also reported a slight overestimation caused by the interface effects, which resulted in deviations of approximately 2.13% and 1.54% near the heterogeneity boundary. For the FFTC method, the mean absolute deviations of depth-dose curves were 1.39% and 1.05% for these two fields, respectively. The maximum deviations observed with the FFTC method were 4.99% and 3.58% near the heterogeneity boundary.

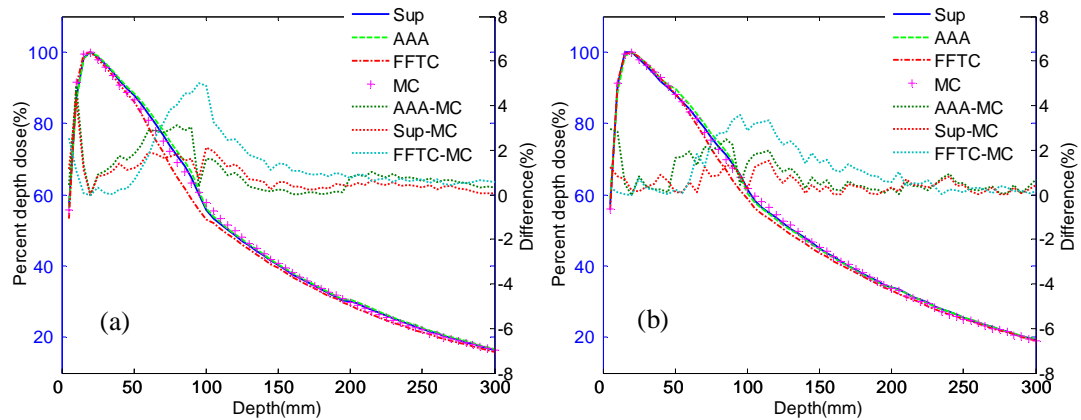


Figure 2.4. Bone slab phantom: comparison of the superposition method (Sup), FFTC method (FFTC), AAA algorithm, and MC simulation. (a) Depth-dose curves of $50 \times 50 \text{mm}^2$ field size; (b) depth-dose curves of $100 \times 100 \text{mm}^2$ field size.

II.6.4. Lung and bone block phantoms

For the lung and bone block phantoms, Fig. 2.5 displays the lateral dose profiles for a $100 \times 100 \text{ mm}^2$ field at a different depth from the phantom surface. The superposition method accurately accounted for attenuations around the lateral material interface for both block phantoms. Mean absolute deviation was 0.94% at 100mm depth in the lung block phantom and 0.57% at 75mm depth in the bone block phantom. The maximum differences for the calculated profiles occurred near the heterogeneity boundary and around the field edges. For these two dose profiles, the largest deviations were 5.93% and 3.95% near the heterogeneity boundary and around the field edges, respectively for the two phantoms, yet remained under 1% in the other regions. Compared with the AAA algorithm, the superposition algorithm exhibited good agreement with the MC simulations, with the accuracy improving from 0.31% to 0.76% in different heterogeneous phantoms (Table 2.1). The FFTC method, which did not take into account the density changes in the lateral direction, could not accurately model the attenuations in the heterogeneities. The result was mean absolute deviations of approximately 1.89% and 1.08% for these two phantoms, respectively.

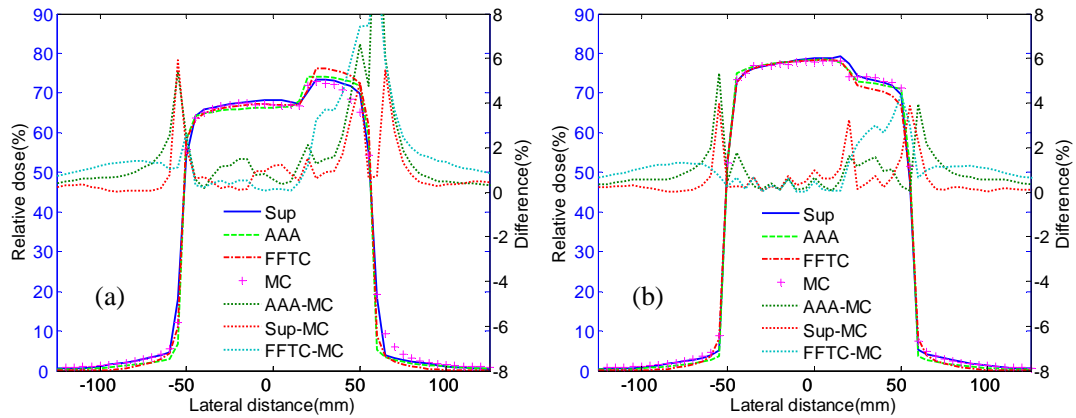


Figure 2.5. Lung and bone block phantoms: comparison of the superposition method (Sup), FFTC method (FFTC), AAA algorithm, and MC simulation for $100 \times 100 \text{ mm}^2$ field size. (a) Profiles at 100mm depth in lung block phantom; (b) profiles at 75mm depth in bone block phantom.

	$100 \times 100 \text{ mm}^2$	$200 \times 200 \text{ mm}^2$
FFTC	4.0 s	4.1 s
Superposition	159.8 s	182.3 s
AAA	170.8 s	184.4 s
MC	12 h	14 h

Table 2.2. Computational time with the FFTC method, superposition method, AAA algorithm, and MC algorithm for $100 \times 100 \text{ mm}^2$ and $200 \times 200 \text{ mm}^2$ field sizes, respectively.

II.6.5. Computation time

The computational times have been presented in Table 2.2. All methods were performed on the same computer (dual-core Intel (R) Core (TM) 2 CPU 6600 platform with 2-GB memory). The run time of the computer implementation of the superposition method came to approximately 159.8 seconds (s) and 182.3s for the 100×100 and 200×200mm² field sizes, respectively. Dose calculations took only 4.0s and 4.1s for 64×64×60 points for these two fields with the FFTC method.

II.7. Discussion

This study has presented a new pencil beam model aimed at addressing the kernel tilting issue and accounting for heterogeneities. The superposition and FFTC methods were employed for dose distribution calculation. The results (Fig. 2.2-2.5) have revealed that the superposition method performed better than the AAA algorithm, achieving mean absolute deviations <1% in comparison with the MC method. Nevertheless, large deviations of up to 6% were detected with this method near the field edges and heterogeneous media boundary (Fig. 2.5) due to the electronic disequilibrium. The superposition method was accelerated by the FFTC technique, which is about 40 times faster (Table 2.2), yet, compared to the MC calculations, the differences observed in interface regions were significant (up to 8%). This proves that the accuracy was markedly decreased.

It should also be noted that a similar strategy for pencil beam dose calculations has been adopted by several authors for treatment planning systems^[25,26]. In their algorithms, the primary beam's oblique incidence was modeled by kernels derived from the beamlets of different incident angles. In contrast, the pencil beam kernels in our proposed model were only modeled once by an MC simulation for the beamlet-normal to surface. The depth-directed components were scaled by moving the entry point's position on the phantom surface and changing the DSS value. As opposed to the method with the AAA algorithm, therefore, the corrections for lateral components were here implemented by using density scaling directly along the spherical shell, without taking density changes in the previous layers into account. More accurate results could be obtained by modeling the beamlet around the central axis in each direction, though this would be more demanding in terms of the memory space required to store the parameters used in dose calculations.

To speed up dose calculations, significant computation time can be saved by performing FFTC calculations on spherical shells without taking into account the density changes in the lateral direction. In heterogeneous tissue, the depth-directed components can be corrected according to the interaction points. As a result, in a spherical coordinate system, the energy released at the same depth is deposited on the same spherical shell. The dose distributions in a phantom can be calculated directly on each spherical shell by means of the FFTC technique. The FFTC technique accelerated the superposition method and saved on computation time, proving approximately 40 times faster than the superposition method (Table 2.2), though accuracy was markedly decreased. Compared

to other FFTC-based methods^[35, 46], our proposed method calculated dose distributions in a spherical coordinate system, avoiding the bias caused by oblique kernels. Improved dose calculation accuracy was also achieved (<3% for mean absolute deviations) compared to other FFTC-based methods reported in the literature (up to 10% for the maximum deviations)^[35, 46, 47]. In IMRT systems, iterative optimization can be completed in a reasonable time by using hybrid dose calculation methods, in which the FFTC-based method is used for most of the dose calculations, with result correction conducted by means of a periodically updated dose correction matrix^[48, 49].

In terms of computational efficiency, the superposition method achieved better calculation speeds than the AAA algorithm, though this effect was very small, as presented in Table 2.2. For hybrid-dose calculation methods in IMRT systems, significant time can be saved by using the FFTC algorithm without reducing optimization quality.

II.8. Conclusion

In this study, we have proposed a new pencil beam model aimed to avoid the bias caused by oblique kernels and to take heterogeneities into account. Our proposed method achieved better dose calculation accuracy without increasing computational time compared to the AAA algorithm. Significant computation time savings can be gained by performing FFTC calculations on spherical shells without taking density changes from the previous layers into account. Nevertheless, the accuracy was greatly decreased with this method (up to 3.0% for mean absolute deviations), as compared to the superposition method. Owing to the lack of cylindrical symmetry for the oblique beamlets, small errors may be introduced into the calculations by correcting the heterogeneous tissue with the symmetrical kernel of different DSSs, yet dose calculations are more accurate with our proposed method. In this study, the spectrum of 6MV photon beams was taken into account in calculating dose distributions in the phantom. Future studies will be focused on modeling the beam for a linear accelerator and conducting dose calculation in patients.

This dose calculation process is not the only crucial part of the Treatment Planning Systems. This calculation is repeated at each iteration of an optimization process during which the plan characteristics (or the fluence map) are refined to minimize a cost function incorporating constraints on both the target and the organs at risk. The incorporation of biological constraints, namely NTCP constraints, is the purpose of the following chapter.

Chapter III. IMRT optimization incorporating equivalent convex NTCP constraint

III.1. Introduction

The primary aim of inverse treatment planning in the setting of intensity modulated radiotherapy (IMRT) is to deliver a therapeutic dose to the target volumes, while minimizing the impact on healthy tissue. Optimized criteria are based either on physical factors or on biological formulations^[8]. The former are often defined as a quadratic dose-based objective function penalizing the mean-square deviations of the various doses and/or dose-volume constraints, as specified in the optimization process. This type of physics-based approach is unable, however, to sufficiently take into account the nonlinear response of tumors or normal structures to irradiation, especially with arbitrary inhomogeneous dose distributions^[50]. Biological indices are capable of reflecting radiobiological responses. In combination with physical criteria, biological criteria-based optimization further improves treatment plan quality. The biological objective function can be employed to improve patient outcome by incorporating both tumor control and normal tissue toxicity probabilities into the inverse planning, as well as striking a finer balance between tumor cure and normal tissue complications. At least three kinds of biological criteria can be applied: (generalized) equivalent uniform dose ((g)EUD), TCP, and NTCP^[51].

The merits of including the (g)EUD concept into an objective function have been widely investigated^[50, 52-56]. Based on a volume parameter a , (g)EUD corresponds, for a given non-uniform dose distribution, to the uniform dose that induces the same biological effect. The (g)EUD model potentially enables treatment plans to be reliably ranked, providing that the value of parameter a is known. Nevertheless, this model cannot directly quantify normal tissue complication probability. In contrast, NTCP models are more clinically relevant, since they estimate, by definition, the toxicity of normal tissue for any given dose distributions. If all constraints of the inverse planning objective function are satisfied, the NTCP-based treatment plan should be implemented, and there is no need to explore other plan options. However, NTCP models require more parameters (typically three), and calibration therefore demands more effort^[57].

The use of NTCP criteria in treatment planning optimization has been investigated by several groups^[58-60], and has already been incorporated into commercial treatment planning software^[57, 60]. Mohan *et al*^[58] used a single score, mixing TCP with NTCP criteria in order to reduce normal tissue complication rates and increase tumor control. The simulated annealing approach was applied to solve this non-convex optimization problem. Wang *et al*^[61] used the same objective function as that proposed by Mohan *et al*^[58], employing the steepest-descent optimization method in order to solve the optimization problem. Stavrev *et al*^[59] designed a hybrid physical-biological model for inverse planning, applying a continuous penalty function method so as to solve the

constrained minimization problem. In this hybrid physical-biological approach, the objective function was calculated as the sum of the weighted NTCP criteria for all organs at risk (OARs), and the dose delivered to the target was controlled by constraints placed on the prescribed minimum and maximum doses. NTCP criteria are sigmoidal functions of dose distributions,^[9] and are thus inherently nonlinear and non-convex in terms of fluence elements in fluence map optimization (FMO). Direct incorporation of such non-linear and non-convex criteria into objective function, as involved in these methods, renders a global optimum extremely difficult to find^[62, 63]. To overcome this limitation, Hoffmann *et al*^[9] and Romeijn *et al*^[64] advanced a theory of how to transform biological criteria into equivalent convex criteria, and discussed the conditions under which these equivalent criteria are strictly convex or concave, depending on the criterion to be minimized or maximized. Their works did not, however, include an investigation of how these equivalent convex criteria could be applied in treatment planning.

This work sought to propose and assess an original method incorporating the equivalent convex NTCP criterion into treatment planning. A hybrid physical-biological method was considered as a solution for the FMO problem. The objective function was defined as the sum of weighted sub-scores for organs under consideration, and each sub-score was composed of physical or biological criteria. The NTCP-based sub-scores were then used to minimize the dose delivered to normal tissues, while dose-based sub-scores were set to achieve the target dose. This constrained optimization problem was then solved by means of a gradient-based optimization algorithm, namely the limited-memory Broyden-Fletcher-Goldfarb-Shanno (L-BFGS) algorithm^[65]. The effectiveness of the proposed method was assessed in 10 prostate cancer cases, and initially compared to the dose-volume based optimization method. Then, the proposed method was compared to the two other NTCP-based optimizations proposed by Mohan *et al*^[58] and Stavrev *et al*^[59].

III.2. Advantages of biological cost functions

III.2.1. Limitations of dose-volume-based treatment planning

Single or multiple dose-volume (DV) constraints used for inverse treatment planning or plan evaluation are based on clinical studies that demonstrate correlation between treatment outcome (e.g., TCP and NTCP) and particular DV metrics^[57]. There are a number of limitations associated with this approach^[57]. (1) Typically, more than one point on the DV histogram (DVH) correlates to the complication, indicating that the different portions of the DVH curve may correlate with risk. This correlation is, however, specific to treatment delivery technique, i.e., intensity-modulated RT (IMRT) or 3D conformal RT (3DCRT), beam arrangements, etc. (2) Generally, optimization with DV constraints is indirect, requiring substantial skill in selecting values and relative weights for constraints to achieve optimal TCP and NTCP values. With typically 1-3 constraints, a range of optimized organ-at-risk (OAR) DVHs that satisfy these few constraints, but carry a distinctly different risk of complications, is possible. (3) Specifying multiple DV constraints increases computational complexity. Moreover, the resulted cost function can

lead to multiple local minima.

III.2.2. Advantages of biological cost functions over dose-volume cost functions

Optimization criteria based on biologically related models are potentially more versatile and directly associated with treatment outcome than those based on DV criteria. Another aspect of plan optimization is that the figure of merit has to address the inevitable variability of patient geometries, and resultant dose distributions, in a population. In this regard, multiple DV criteria for a single organ quickly may become problematic as they need to be given an individual priority and ideally ought to be combined into a single figure of merit to avoid ambiguities. In contrast, biologically related models have the potential to provide an inherent prioritization of multiple DV criteria incorporated in a single figure of merit.

III.3. Methods and Materials

III.3.1. Dose calculations

The standard pencil beam model proposed by Ahnesjö^[29] was applied to calculate dose distributions. The dose contribution of each beam element to each voxel for unit fluence was pre-computed and recorded in a matrix. The dose distributions could be expressed in a matrix-vector form as:

$$\mathbf{D}(\mathbf{x}) = \mathbf{W}\mathbf{x}, \quad (3.1)$$

where $\mathbf{D}(\mathbf{x})$ is the dose distribution vector (dose for each patient voxel), \mathbf{W} is the dose calculation matrix, and \mathbf{x} is the fluence elements (beamlet weights).

III.3.2. Equivalent convex NTCP constraint

In this study, we investigated with a hybrid physical-biological method to estimate the inhomogeneous dose response for tumors and normal tissues. The high-dose delivered to a target was constrained by physical criteria, such as quadratic dose-based sub-scores. The risks of normal tissue complications were controlled by NTCP criteria. The NTCP model we applied was the LKB model^[66, 67], defined by:

$$NTCP_{LKB}[\mathbf{D}(\mathbf{x})] = \Phi\left(\frac{gEUD(\mathbf{D}(\mathbf{x})) - D_{50}}{mD_{50}}\right), \quad (3.2)$$

where $\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^z e^{-t^2/2} dt$ is the standard normal cumulative distribution function; D_{50} represents the tolerance dose to a whole organ, causing a 50% complication probability; m is the slope of the sigmoidal function Φ ; and $gEUD_a(\mathbf{D}(\mathbf{x}))$ is the generalized equivalent uniform dose^[68] of the dose distribution $\mathbf{D}(\mathbf{x})$, given by:

$$gEUD[\mathbf{D}(\mathbf{x})] = \left(\frac{1}{N_0} \sum_{i=1}^{N_0} [\mathbf{D}_i(\mathbf{x})]^a \right)^{\frac{1}{a}}. \quad (3.3)$$

N_0 is the number of voxels in the anatomic structure of interest, \mathbf{D}_i is the dose in the i th voxel, and a is the tissue-dependence parameter that represents the dose-volume effect. A large a value indicates serial organ architecture that is strongly sensitive to high dose, whereas a small value (a near to 1) indicates parallel organ architecture that is sensitive to the mean dose.

Radiobiological criteria like TCP and NTCP are often sigmoid functions of dose distributions^[9, 64], such as the LKB model in Eq. (3.2). The NTCP model is thus a non-linear and non-convex function in terms of fluence elements. In order to ensure that the optimization problem can be easily solved using a gradient-based method, the NTCP-based sub-score was calculated using the following equivalent convex NTCP criterion^[9]

$$-\ln(1 - NTCP_{LKB}(\mathbf{D}(\mathbf{x}))). \quad (3.4)$$

For this NTCP model, the corresponding sub-score was penalized when the NTCP value was superior to the desired value of $NTCP_0$. It can be written as follows:

$$f_{NTCP}(\mathbf{D}(\mathbf{x})) = H \left(\ln \frac{1 - NTCP_0}{1 - NTCP_{LKB}(\mathbf{D}(\mathbf{x}))} \right) \ln \frac{1 - NTCP_0}{1 - NTCP_{LKB}(\mathbf{D}(\mathbf{x}))} \quad (3.5)$$

where H is a step function.

III.3.3. Hybrid physical-biological model

By using fluence map optimizations, we aimed to find a solution within the search space for the inverse treatment planning. Given the matrix \mathbf{W} , our objective consisted of finding a suitable vector \mathbf{x} that would satisfy the optimization goals. The objective function f that was used in this work was defined as the sum of weighted sub-scores f_l for all organs under consideration

$$f(\mathbf{D}(\mathbf{x})) = \sum_{l=1}^L \xi_l f_l(\mathbf{D}(\mathbf{x})), \quad (3.6)$$

where the weights ξ_l represent the differences in clinical significance and L is the number of sub-scores. It is worth noting that more than one sub-score can be applied to the same organ in order to improve dose distributions.

The objective function defined in Eq. (3.6) was applied for inverse planning of prostate cancer cases. This consisted specifically of applying a minimum dose sub-score of $f_{Min}(\mathbf{D}_{PTV}(\mathbf{x}))$ and a mean dose sub-score of $f_{Mean}(\mathbf{D}_{PTV}(\mathbf{x}))$ to guarantee the therapeutic dose delivered to the PTV (planning target volume), along with two NTCP-based sub-scores of $f_{NTCP}(\mathbf{D}_{Bladder}(\mathbf{x}))$ and $f_{NTCP}(\mathbf{D}_{Rectum}(\mathbf{x}))$ to minimize the dose delivered to the bladder and rectum walls, respectively. In order to achieve a fast decrease of dose in the region outside the PTV, a dose-volume sub-score of $f_{DVH}(\mathbf{D}_{Healthy}(\mathbf{x}))$ was added to other healthy tissue areas. This scheme can be mathematically expressed as

$$f(\mathbf{D}(\mathbf{x})) = \xi_1 f_{NTCP}(\mathbf{D}_{Bladder}(\mathbf{x})) + \xi_2 f_{NTCP}(\mathbf{D}_{Rectum}(\mathbf{x})) + \xi_3 f_{Min}(\mathbf{D}_{PTV}(\mathbf{x})) + \xi_4 f_{Mean}(\mathbf{D}_{PTV}(\mathbf{x})) + \xi_5 f_{DVH}(\mathbf{D}_{Healthy}(\mathbf{x})) \quad (3.7)$$

In this work, we applied the convex dose-based sub-scores of $f_{Min}(\mathbf{D})$, $f_{Mean}(\mathbf{D})$, and $f_{DVH}(\mathbf{D})$ proposed by Wu and Mohan^[69].

III.3.4. Optimization method of the fluence elements

Figure 3.1 displays the flowchart for implementing the proposed algorithm. We employed an iterative gradient technique to optimize the fluence elements. The acceptable step length in the search direction was computed by means of a normal line search method applying the Wolfe conditions, a set of inequalities for performing inexact line searches. The inverse Hessian matrix was updated by means of the L-BFGS algorithm^[65]. Our proposed method required the first derivatives of sub-scores in order to compute the step length and update the Hessian matrix. For the NTCP-based sub-score, its first derivative could be calculated with the following formula:

$$\frac{\partial f_{NTCP}}{\partial \mathbf{x}_j} = H \left(\ln \frac{1 - NTCP_0}{1 - NTCP_{LKB}(\mathbf{D}(\mathbf{x}))} \right) \frac{1}{1 - NTCP_{LKB}(\mathbf{D}(\mathbf{x}))} \frac{1}{\sqrt{2\pi m D_{50}}} e^{-t^2} \frac{gEUD(\mathbf{D}(\mathbf{x}))^{N_0}}{\sum_{l=1}^{N_0} (\mathbf{D}_l(\mathbf{x}))^a} \sum_{i=1}^{N_0} (\mathbf{D}_i(\mathbf{x}))^{a-1} \mathbf{W}_{i,j}, \quad (3.8)$$

where $t = \frac{gEUD(\mathbf{D}(\mathbf{x})) - D_{50}}{\sqrt{2m D_{50}}}$, and $\mathbf{W}_{i,j}$ is the (i, j) th element of the matrix \mathbf{W} .

For the dose-based sub-scores, the first derivatives were provided as follows:

i) mean dose sub-score

$$\frac{\partial f_{Mean}}{\partial \mathbf{x}_j} = \frac{2}{N_0} \sum_{i=1}^{N_0} [\mathbf{D}_i(\mathbf{x}) - D_0] \mathbf{W}_{i,j}, \quad (3.9)$$

where D_0 is the prescribed dose.

ii) minimum dose sub-score

$$\frac{\partial f_{Min}}{\partial \mathbf{x}_j} = \frac{2}{N_0} \sum_{i=1}^{N_0} H[D_0 - \mathbf{D}_i(\mathbf{x})] \cdot [\mathbf{D}_i(\mathbf{x}) - D_0] \mathbf{W}_{i,j}, \quad (3.10)$$

iii) dose-volume sub-score

$$\frac{\partial f_{DVH}}{\partial \mathbf{x}_j} = \frac{2}{N_0} \sum_{i=1}^{N_0} H[\mathbf{D}_i(\mathbf{x}) - D_1] \cdot H[D_2 - \mathbf{D}_i(\mathbf{x})] \cdot [\mathbf{D}_i(\mathbf{x}) - D_1] \mathbf{W}_{i,j}, \quad (3.11)$$

where the values of D_1 and D_2 were calculated based on the prescribed dose-volume constraint ($V_{D1} < V_1$), *i.e.* the volume of the organ receiving a dose greater than D_1 should be less than V_1 .

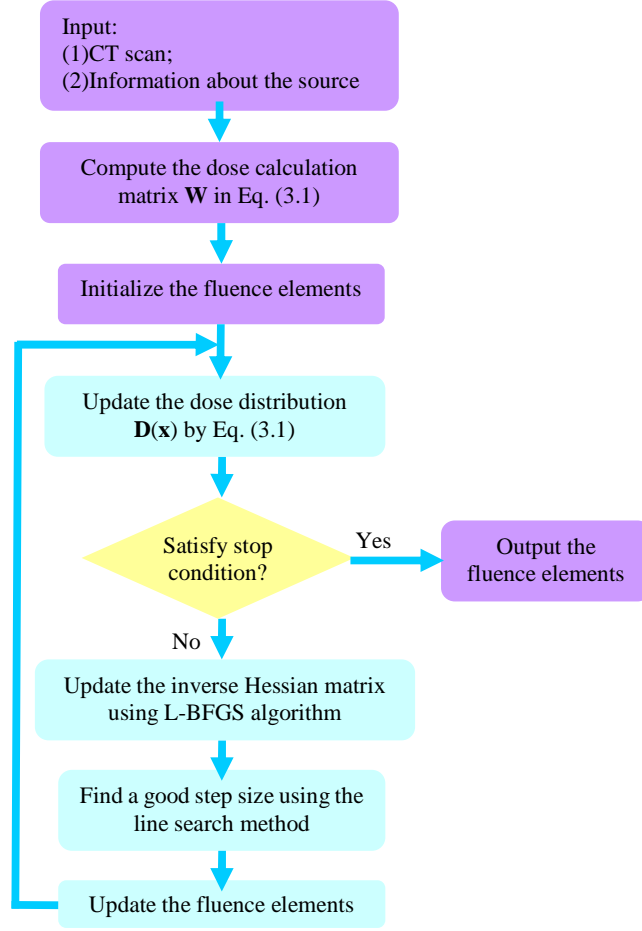


Figure 3.1. Flowchart of the method with convex optimization.

III.3.5. NTCP-based optimizations proposed by Mohan *et al* and Stavrev *et al*

Our method was compared to two other NTCP-based optimizations reported by Mohan *et al* ^[58] and Stavrev *et al* ^[59], respectively. Mohan *et al*'s proposed method consisted of using a single score, with mixed TCP and NTCP criteria, to reduce normal tissue complication and increase tumor control ^[58]. In their study, two NTCP-based sub-scores were applied to the rectum and bladder walls, respectively, with one TCP-based sub-score set to the target. In contrast, for the Stavrev *et al*'s proposed optimization method, a hybrid physical-biological model was designed for inverse planning ^[59]. The objective function comprised two NTCP-based sub-scores (for the rectum and bladder walls, respectively), and the dose delivered to the target was controlled by a selection of constraints (see reference ^[59] for more details). Both of these methods applied the LKB NTCP model. The TCP and NTCP parameters used in this work were identical to those given in reference ^[58, 59]. Corresponding optimization problems were solved by means of a simulated annealing (SA) and nonlinear programming (NP) method, respectively. These two optimizations were labeled as 'SA' and 'NP', according to the optimization algorithm used. Our proposed method was labeled as 'EC-NTCP', representing the use of the

equivalent convex NTCP criteria.

Our method was also compared to the dose-volume-based optimization approach, which applied the convex dose-based sub-scores, as mentioned in Sec. III.2.3. The dose-volume-based optimization was labeled as ‘DVH’. The objective functions for both our proposed method and the dose-volume-based optimization were the convex functions. These two optimization methods can therefore be solved by the same gradient-based optimization algorithm, namely the L-BFGS. The optimization procedure was interrupted when the score value (Eq. (3.7)) was sufficiently decreased ($<1e-6$) or the line search was unable to locate an acceptable point along the current search. For the NTCP-based optimizations proposed by Mohan *et al* ^[58] and Stavrev *et al* ^[59], the optimization algorithm displayed in figure 1 should be replaced by a simulated annealing and nonlinear programming method, respectively.

III.3.6. Organ delineation and dose prescription for 10 patients

Our proposed method was tested in 10 prostate cancer cases. The patients underwent simulation and treatment in the supine position. Target volume and organs at risk (bladder, rectum, and femoral heads) were delineated on CT slices. The rectum and bladder walls were generated with a thickness of 5mm from the external manually-delineated rectal and bladder contours, respectively. A “tissue ring”, the outside of the area extending the PTV by 5cm, was defined as the other normal tissue used in Eq. (3.7). The target volume included the prostate and excluded the pelvic lymph nodes. The PTV was calculated by adding a 10mm margin in all directions except the posterior, where a 5mm margin was applied. Our method used five coplanar 6MV photon beams, with gantry angles of 36°, 100°, 180°, 260°, and 324°, respectively. The dose matrix \mathbf{W} was calculated by applying the standard pencil beam algorithm ^[29], implemented into the CERR software ^[70]. The plans were normalized so as to deliver the 78Gy prescription dose to the prostate (PTV). In order to produce the highly homogeneous target dose coverage, both minimum and mean dose sub-scores (in Eq. (3.7)) were used to control the dose distributions in the target. The NTCP radiobiological parameters for the bladder and rectum originated from the works of Dale *et al* ^[71] and Peeters *et al* ^[72] (see Table 3.1), respectively. In this work, both the prescribed bladder and rectum toxicities were set to zero.

Table 3.1. The radiobiological parameters of NTCP model for various organs used in the treatment planning optimization.

Organs	$n(1/a)$	m	$D_{50}(\text{Gy})$	Endpoint
Bladder ^[71]	0.13	0.11	62	RTOG grade ≥ 3 bladder complications
Rectum ^[72]	0.13	0.14	80.7	Severe proctitis/necrosis/stenosis/fistula

III.3.7. Software for optimization and comparison test

The SIMULANNEALBND MATLAB code was applied to perform the simulated annealing method by replacing its generator function with a Gaussian generator ^[73]. The FMINCON MATLAB code was applied for the nonlinear programming method. To assess

the relative quality of each treatment plan, optimization results were evaluated according to practical clinical guidelines, as specified by Marks *et al* ^[74], which are outlined in Table 3.2. The different optimization methods were compared in terms of the following parameters: gEUD, NTCP (in Eq. (3.2)), TCP, number of iterations, computation time, and score values. A non-parametric test (Wilcoxon test) was used for significance level comparison.

Organ		Dose/volume parameters			
Bladder		$V_{65}<50\%$	$V_{70}<35\%$	$V_{75}<25\%$	$V_{80}<15\%$
Rectum	$V_{50}<50\%$	$V_{60}<35\%$	$V_{65}<25\%$	$V_{70}<20\%$	$V_{75}<15\%$

Table 3.2. Dose-volume constraints in the OARs for physical optimization.

III.4. Results

III.4.1. Comparison of the proposed method with dose-volume-based optimization

	Bladder	Rectum	PTV		Tissue ring
	NTCP	NTCP	Minimum dose	Mean dose	DVH
Patient 1	60	90	30	30	5
Patient 2	45	100	30	30	5
Patient 3	45	80	30	30	5
Patient 4	45	110	30	30	5
Patient 5	50	100	30	30	5
Patient 6	120	70	30	30	5
Patient 7	45	80	30	30	5
Patient 8	45	60	30	30	5
Patient 9	100	120	30	30	5
Patient 10	60	50	30	30	5

Table 3.3. The weights of the proposed optimization (in Eq. (7)) in 10 different patients.

Our method was compared to dose-volume-based optimization in 10 prostate cancer cases. To that end, we used both DVH curves and NTCP values as endpoints. In our method, only one NTCP-based sub-score was applied to minimize the dose delivered to one OAR. For dose-volume-based optimization, the dose distribution in one OAR was controlled by replacing one NTCP-based sub-score of $f_{\text{NTCP}}(\mathbf{D}(\mathbf{x}))$ in Eq. (3.7) with three dose-volume sub-scores. Some sections of the OARs (rectum and bladder walls) were covered by the PTV. Decreasing the high-dose of OARs led to a decrease in the dose delivered to PTV. In order to achieve a similar DVH curves for two different optimization methods, the weights of two objective functions were determined by trial and error. The weights were adjusted for the same patient in order to attain similar tumor dose coverage and OAR sparing in the different optimization methods. The weights chosen for both the proposed method and multi-dose-volume-based optimization in the 10 patients are provided in Tables 3.3 and 3.4, respectively, and the dosimetric results in Table 3.5. The dose distributions in PTV and OARs were quantified based on mean dose and NTCP

values, respectively. Figure 3.2 displays the comparison between the average DVH curves of 10 prostate cancer cases of both approaches. In comparison with dose-volume-based optimization, the proposed method has the advantage in terms of high-dose control (see Fig. 3.2), even though it results in a slight decrease in control for the low-dose (<50Gy) region. Nevertheless, the impact on this dose region is limited. According to the practical clinical guidelines outlined by Marks *et al* ^[74], the dose-volume constraint is not considered for doses of less than 50Gy in the rectum and bladder walls.

	Bladder			Rectum		PTV		Tissue ring	
	DVH	DVH	DVH	DVH	DVH	DVH	Minimum dose	Mean dose	DVH
Patient 1	5	5	5	5	8	30	180	180	5
Patient 2	5	5	5	5	8	20	180	180	5
Patient 3	5	5	5	8	15	15	180	180	5
Patient 4	8	8	8	12	8	10	180	180	5
Patient 5	3	3	3	15	15	11	180	180	5
Patient 6	50	50	50	8	12	10	180	180	5
Patient 7	1	1	1	8	8	8	180	180	5
Patient 8	10	10	10	10	10	10	180	180	5
Patient 9	10	1	1	10	10	10	180	180	5
Patient 10	10	15	1	10	5	5	180	180	5

Table 3.4. The weights of the optimization based on multi dose-volume sub-scores in 10 different patients.

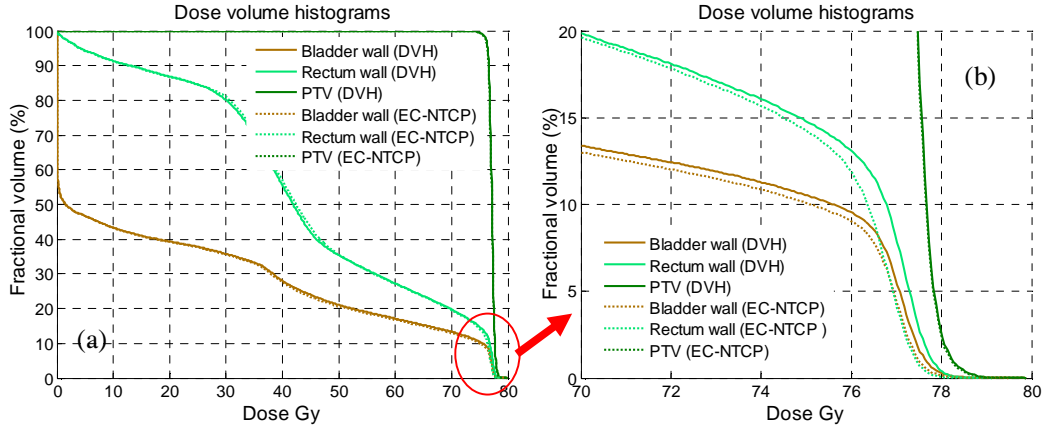


Figure 3.2. The average dose-volume histograms of 10 patients for rectum wall, bladder wall, and PTV in prostate cancer cases. The solid lines represent the DVH curves of the dose-volume based optimization, and the dotted lines the DVH curves of the proposed method (EC-NTCP). (a) DVH curves of these two method types; (b) high dose region of DVH curves.

As illustrated in Fig. 3.3, we calculated p values (Wilcoxon test) for the different dose bins of dose-volume histograms from Fig. 3.2. Significant differences were observed between the proposed method results and those of the dose-volume-based method for the bladder, at between 20Gy and 75Gy ($p < 0.05$), as demonstrated in Fig. 3.3(a). For the rectum (see Fig. 3.3(b)), significant differences were observed in the medium (30-50Gy)

and high dose regions (75-78Gy). PTV coverage was maintained for the different optimization methods. We then compared these different methods using the outputs of normal tissue (bladder and rectum). The high p values for PTV represent the same dose coverage in PTV for different optimizations (Fig. 3.3(c)).

	Bladder		Rectum		PTV	
	DVH	EC-NTCP	DVH	EC-NTCP	DVH	EC-NTCP
patient 1	18.49%	16.87%	6.02%	5.80%	77.04	77.03
patient 2	71.45%	70.41%	5.08%	4.85%	77.08	77.07
patient 3	41.47%	40.03%	12.45%	11.64%	77.13	77.09
patient 4	55.77%	54.47%	6.32%	6.07%	77.22	77.20
patient 5	89.32%	88.71%	5.93%	5.63%	77.31	77.28
patient 6	79.31%	77.72%	4.66%	4.35%	77.05	77.00
patient 7	3.75%	3.08%	4.9%	4.71%	77.06	77.04
patient 8	28.42%	27.4%	4.34%	4.08%	77.04	77.04
patient 9	68.06%	66.75%	7.00%	6.84%	77.13	77.08
patient 10	62.83%	61.04%	4.54%	4.56%	76.92	76.90
Overall	51.89%	50.65%	6.12%	5.85%	77.10	77.07

Table 3.5. The dosimetric results for the 10 patients. NTCP values for bladder and rectum, and mean dose (Gy) for PTV.

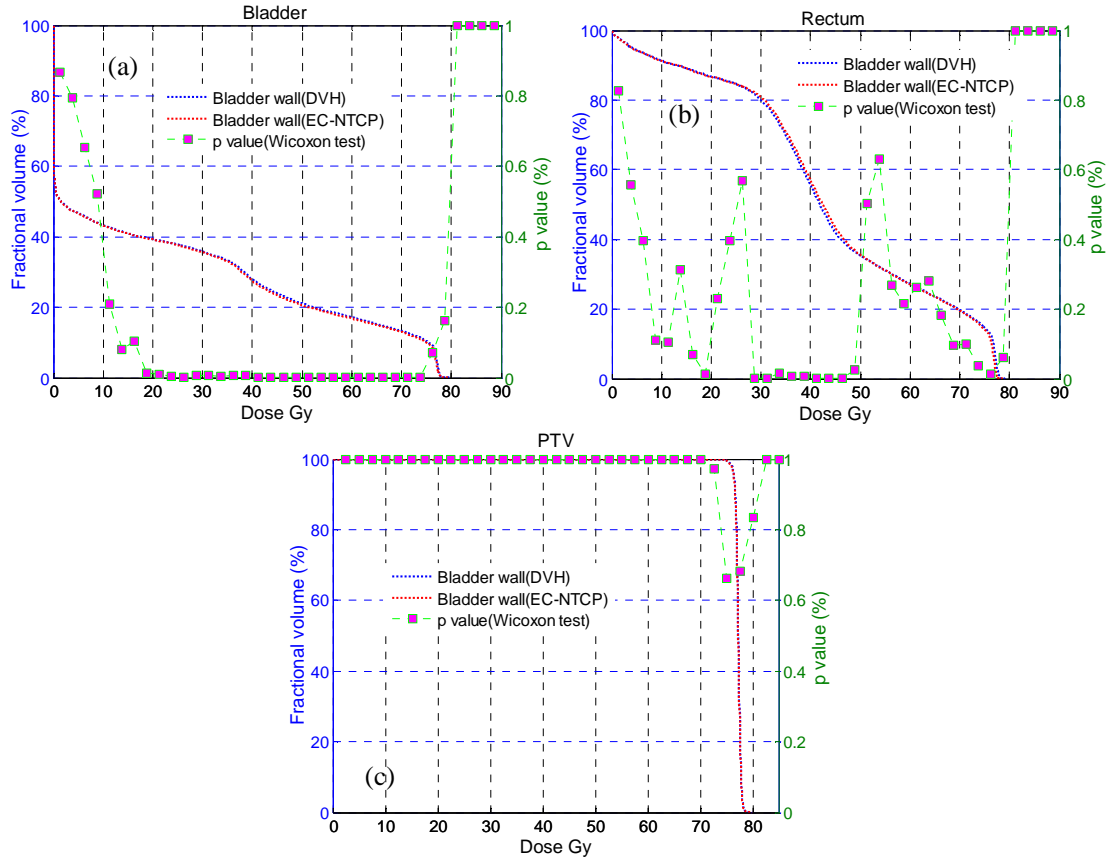


Figure 3.3. Significant differences between the dose-volume-based optimization and the proposed method for: (a) bladder wall; (b) rectum wall; (c) PTV.

III.4.2. Comparison of the proposed method with SA and NP

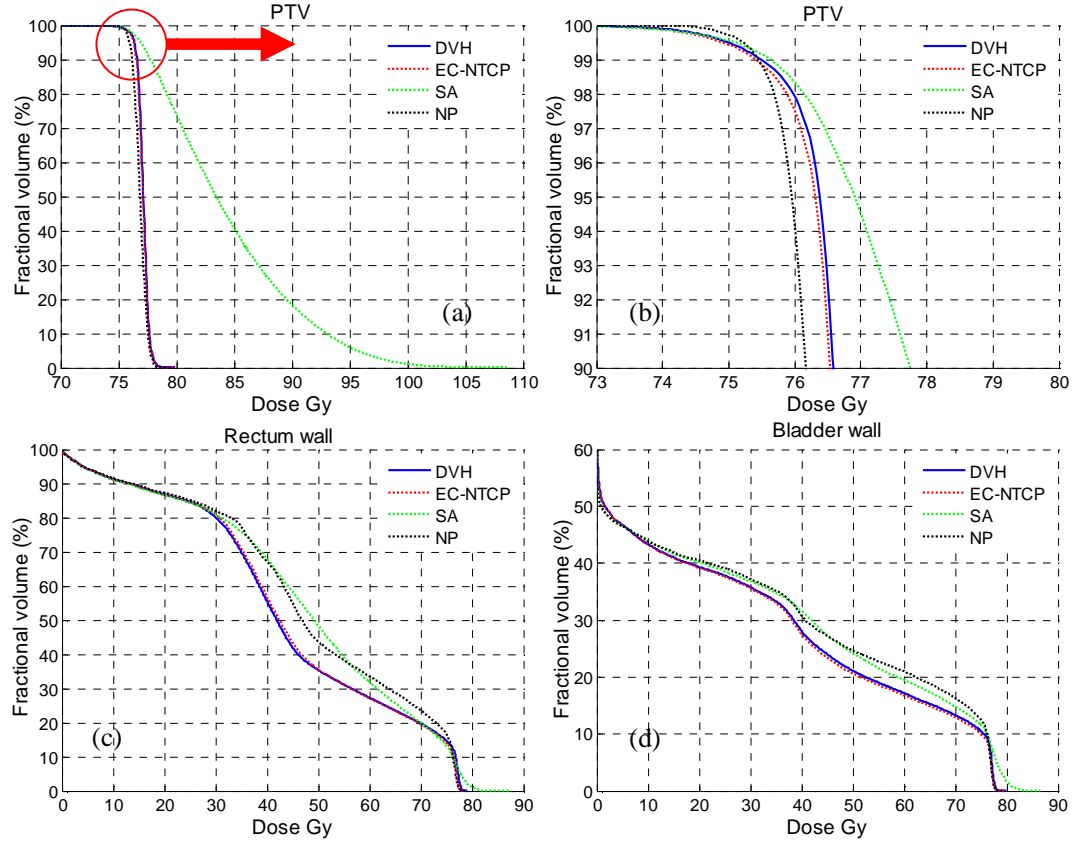


Figure 3.4. The average dose-volume histogram of 10 patients for the plan with physical criteria (DVH), proposed method (EC-NTCP), plan optimized by simulated annealing method (SA), and plan optimized by nonlinear programming method (NP). DVH curves of: (a) PTV; (b) low dose region of PTV (in the red circle of figure 6 (a)); (c) rectum wall; (d) bladder wall.

Our method was compared to the optimization methods proposed by Mohan *et al* ^[58] and Stavrev *et al* ^[59], respectively, in 10 prostate cancer cases, with the results displayed in Fig. 3.4. For the plan based on Mohan *et al.*'s proposition ^[58], three weights were applied to these three sub-scores, respectively, in order to achieve the best tumor coverage and OAR sparing balance in each patient. As for the Stavrev *et al.*'s optimization approach ^[59], two weights were chosen for the two NTCP sub-scores in the objective function, respectively. In this plan, the dose delivered to the target was controlled by a 76Gy minimum dose, and a 79Gy maximum dose. The same setup was implemented for the proposed method and the optimization based on three dose-volume-based sub-scores, as mentioned before in section III.3.2. Figure 3.4 displays the DVH comparison between our method and the optimizations proposed by Mohan *et al* ^[58] and Stavrev *et al* ^[59]. In comparison with our method, the optimization approach proposed by Mohan *et al* ^[58] exhibited less control of the high-dose distribution in the PTV (Fig. 3.4(a)), and the control of high-doses in OAR was poor due to parts of the bladder and rectum walls being located in the PTV (Fig. 3.4 (c) and (d)). The results obtained with Stavrev *et al.*'s method ^[59] exhibited better control for

the PTV and worse control for the OAR in comparison with our method (Fig. 3.4 (b) and (c)). Yet the improvement provided by this optimization method for PTV coverage is achieved at the expense of an increased NTCP value (in Eq. (3.2)), with around 3.75% for the bladder wall and 1.28% for the rectum wall. The dosimetric results of our method in the 10 patients revealed the lowest gEUD and NTCP values for both the bladder and rectum walls, as shown in Table 3.6. We calculated the gEUD values (in Eq. (3)) and TCP values for the PTV of 10 patients, as outlined in Table 3.6. The parameter $a=-10$ was applied in the gEUD calculation of PTV dose distributions. We used the same TCP model and parameters as those implemented by Mohan *et al* ^[58] to calculate the TCP values. With the exception of Mohan *et al*'s method ^[58], the other three optimization methods were revealed to be capable of achieving similar tumor dose coverage, as shown in Table 3.6. We also computed the number of iterations, computation time, and last score values, all of which can be found in Table 3.7. All these methods were performed using the same computer (Intel(R) Xeon(R) 4CPU E5430 platform with 4GB memory).

	Bladder		Rectum		PTV	
	gEUD (Gy)	NTCP	gEUD (Gy)	NTCP	gEUD (Gy)	TCP
DVH	62.07	51.89%	63.03	6.12%	77.06	53.77%
EC-NTCP	61.76	50.65%	62.78	5.85%	77.04	53.70%
SA	63.27	57.08%	63.70	7.08%	82.67	68.44%
NP	62.57	54.40%	63.98	7.13%	76.80	53.10%

Table 3.6. Average gEUD values and NTCP values in 10 prostate cases for bladder, rectum, and TCP values for PTV for the plans obtained with physical criteria (DVH), proposed optimization (EC-NTCP), optimized by simulated annealing method (SA), and optimized by nonlinear programming method (NP).

	DVH	EC-NTCP	SA	NP
No. iterations	39	44	10000	99
Time (s)	96.27	98.66	166.43	799.63
Score values	0.1146	0.1064	0.3971	0.1409

Table 3.7. Number of iterations, computation times, and last score values for the plan with physical criteria (DVH), proposed method (EC-NTCP), plan optimized by simulated annealing method (SA), and plan optimized by nonlinear programming method (NP).

III.5. Discussion

In this study, we proposed a hybrid physical-biological model for inverse planning optimization, consisting of controlling the dose delivered to normal tissue by NTCP-based sub-scores and achieving the desired target dose by dose-based sub-scores. This large-scale constrained optimization problem was then solved by means of the L-BFGS algorithm ^[65]. The proposed method was tested in 10 prostate cancer cases. The results proved that the proposed method was capable of generating clinically acceptable plans which meet the practical clinical guidelines ^[74].

To prove the applicability of the proposed method, experiments were conducted with comparisons of the dose-volume-based optimization method. The average results of the 10 patients (Fig. 3.4) were similar for the two different optimization methods. The DVH based scheme can generate the optimal results. For example, if the DVH curves of the optimal result are known, we can set the DVH constraints in the objective function according to the control point on the DVH curves and get the similar result as the optimal result. However, in the practical application, the optimal result is unknown. So, we cannot set the appropriate DVH constraints in the objective function at once. In the clinic application, we have to try several times to get the “optimal” result. In these tries, the DVH constraints are adjusted appropriately to protect the normal tissue and increase the tumor control. This “optimal” result is the best result in the results we have tried. We cannot guarantee this “optimal” result is the optimal result for the given patient, because no NTCP or TCP criterion is applied to evaluate the calculated result in the optimization. This is one drawback of the application for the DVH based scheme.

Some papers have reported the practical clinical guidelines, such as QUANTEC^[74]. The DVH constraints can be obtained from these practical clinical guidelines. The practical clinical guidelines are obtained from the statistics of a large number patient. For a given patient, the practical clinical guidelines may be not suitable. For example, the $V_{70} \leq 35\%$ is reported by Marks *et al*^[74] for the bladder wall. For a patient, a large part of the bladder wall is included in the PTV. Setting the DVH constrain ($V_{70} \leq 35\%$) for bladder wall in the objective function, the DVH based scheme maybe not find a result meeting the current constrain. In contrast, for a patient, a little or no part of the bladder wall is covered by the PTV. Using the DVH based scheme, we can easily find a result in which the V_{70} for the bladder is far less than 35%. Thus, we cannot get the optimal result using the DVH constrain ($V_{70} \leq 35\%$) for bladder wall in the objective function. In the clinical application, the DVH constraints originated from the practical clinical guidelines should be adjusted appropriately to generate the optimal patients for the given patient. This is another drawback of the application for the DVH based scheme.

Overall, both the DVH based scheme and NTCP based scheme can generate the optimal IMRT plan. But, compared to the NTCP based scheme, the DVH based scheme has to try several times to get the “optimal” results. In our study, we used the NTCP based scheme to get the clinically acceptable plans, which meet the practical clinical guidelines^[74]. These results were also compared to the plans generated by the DVH based scheme. The calculated results were similar for the two different optimization methods, because both schemes generate the optimal IMRT plan. As shown in Fig. 3.2 and 3.4, there is the small gain in NTCP based scheme. This small gain is caused by the slight drop in target coverage. The difference between the results gotten from two optimization methods could be narrowed by adding more DVH constraints in the objective functions. In our study, we only used three DVH constraints to control the dose distributions in one normal tissue. The results meet the practical clinical guidelines^[74]. The purpose of our study is to prove that the proposed NTCP schemes can generate clinically acceptable plans. Hence, we did not add more DVH constraints to one organ.

The biological significance of NTCP could be lost by randomly selecting parameters.

Yet numerous papers have investigated the choice of parameters for NTCP models ^[57, 71, 72, 75-81]. In this study, we applied the different sets of parameters reported by Fiorino *et al* ^[77] in the optimizations for 10 prostate cancer cases and recorded the NTCP values. We used the same parameters as Fellin *et al* ^[78], Söhn *et al* ^[79], and Tucker *et al* ^[80] for the RTOG Grade 2-3 rectum, and those applied by Peeters *et al* ^[72] and Rancati *et al* ^[81] for RTOG Grade 3 rectum. The Peeters *et al*'s ^[72] parameters were chosen for the rectum in the experiments (Sec. III.3). We performed the Wilcoxon test for the NTCP values, which were obtained through application of the parameters reported by the other authors, along with the results obtained by Peeters *et al*'s ^[72] parameters. The results revealed that our proposed method is suitable for NTCP models with different parameters for all p values inferior to 0.0046.

We also compared our method with the two other NTCP-based optimization approaches, as reported in ^[58, 59] in reference. Figure 3.4 illustrates that our method achieved higher treatment quality than the two other approaches, with Mohan *et al*'s ^[58] method providing less control of the high-dose distribution in the PTV, as displayed in Fig. 3.4(a). There could therefore be a reduction in control of the high-dose in OAR due to areas of normal tissue being included in the PTV (Fig. 3.4(c) and (d)). Stavrev *et al*'s ^[59] method demonstrated better control for PTV, yet worse for OAR. As regards the rectum wall, the NTCP values (Table 3.6) of the plan based on Mohan *et al*'s method were better than those obtained with that of Stavrev *et al* ^[59]. This value is the result of a compromise between reducing large parts of low dose and increasing small parts of high dose. The rectum wall constitutes a serial organ, sensitive to high doses. Different optimization algorithms were applied in order to solve the optimization problem, according to given forms of objective function. The results revealed that the gradient-based method used in our method exhibited higher convergence efficiency and required less iterations and computation time than the simulated annealing and nonlinear programming methods (Table 3.7). With respect to time consumption, the Wilcoxon test was performed for our proposed method compared to the simulated annealing ($p < 0.014$) and nonlinear programming ($p < 0.002$) methods, respectively. This demonstrated that time saving was a proven advantage of the gradient-based method.

III.6. Conclusion

In this work, we proposed a method incorporating the equivalent convex NTCP constraints into an optimization approach. We described the sub-score of an equivalent convex NTCP model ^[9], and provided its first derivatives. The efficacy of the equivalent convex NTCP constraints, along with their use in large-scale constrained optimization, was investigated. A gradient-based optimization algorithm, namely L-BFGS, was applied so as to solve this optimization problem. The proposed method was compared to dose-volume-based optimization approaches in 10 prostate cancer cases. Our method was proven capable of generating clinically acceptable plans. Furthermore, NTCP-based optimization appeared to be more relevant in terms of clinical interpretation, while providing the most favorable overall treatment outcome. The proposed method was also compared to two other NTCP-based optimizations, solved by simulated annealing and

nonlinear programming methods, respectively. This comparative study showed that the proposed scheme achieved higher quality treatment plans and required less computation time than the two other optimization approaches. Future investigation will address the extension of the proposed physical-biological model to applications in direct aperture optimization.

Combined with the dose calculation method proposed in the previous chapter, this optimization method enables to optimize the plan according to the anatomy of the patient as observed in the planning CT. However, some geometrical uncertainties occur during the treatment and may be compensated for at the planning step. This is the goal of the following chapter.

Chapter IV. The corrections of random geometric uncertainties in IMRT

IV.1. Introduction

Conventional methods of adding a margin around the clinical target volume (CTV) so as to obtain a planning target volume (PTV), aim to find a sound compromise between maximizing target dose and minimizing dose to OAR^[82]. Several methods have been proposed to account for geometric uncertainties in dose calculation, such as the “stochastic simulation method”^[83], the “dose convolution method”^[84, 85], and the “fluence modification method”^[86]. However, these methods inevitably expand CTV area. To significantly reduce the PTV margin, other authors^[17, 87-90] have used the “robust beam profile” method, of which the most important feature is the higher fluence delivered on the field edges.

In this part of work, we concentrated our attention on random geometrical uncertainties. Various deconvolution algorithms have been investigated in order to calculate robust beam profile with respect to small random geometrical uncertainties^[17, 87], and to remove the finite detector size effect in the measured profiles^[91, 92]. To avoid infeasible or unreliable deconvolution results using the direct inverse filtering in the frequency domain, the series expansion method has been employed for first-order approximation of the deconvolution^[17, 91-94]. High-frequency components were suppressed through the summation of the first four terms of series expansion. The series expansion method can effectively suppress the oscillations in deconvolution results, but the effect is limited, and the convolved-back results inevitably contain numerous high frequency components, especially around the field edges. The method proposed by Fan and Nath^[17] is one kind of series expansion method that suppresses the fluence around the field edges. If the suppression is too large, however, it could lead to inferior intensity in the convolved-back profiles than in the nominal static fluence map around the field edges. Here we proposed a new deconvolution method based on series expansion and a Butterworth filter. In order to suppress the high-frequency components, we used only the first four terms of the series. By adjusting the parameters of the Butterworth filter according to the different probability density functions (PDF), we further suppressed high-frequency signal components and minimized differences between the static and convolved-back profiles, especially on the field edges.

This part of work is organized as follows. First, the methods and principles used in the design of the algorithm are described. Then, the new deconvolution approach is detailed (Section IV.3). The results are presented in Section IV.4, including a comparison to the deconvolution kernel method on a regular 2D fluence map, a real IMRT field and a prostate case.

IV.2. Methods and Materials

Geometrical uncertainty results from many independent random causes (setup variation, motion, equipment precision limit, etc.), as discussed above. According to the central limit theorem, the distribution of this uncertainty should converge to the Gaussian distribution in a multi-fraction treatment^[17]. We used a single Gaussian distribution to describe random patient motion, as reported before on robust fluence^[17, 95]:

$$p(\sigma, x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{x^2}{2\sigma^2}\right). \quad (4.1)$$

Here, the standard derivation σ describes the random motion. Given a beam profile, the convolved-back profile $\mathbf{D}_m(x)$ is defined as:

$$\mathbf{D}_m(x) = \int p(\sigma, x-u) \mathbf{D}_p(u) du, \quad (4.2)$$

where $\mathbf{D}_p(u)$ is the deconvolution profile. Our goal was to obtain the optimal deconvolution fluence map $\mathbf{D}_p(u)$ by minimizing the differences between the convolution-back fluence map $\mathbf{D}_m(x)$ and the nominal static fluence map.

IV.2.1. The beam profile deconvolution with the series expansion method

With the Taylor serial expansion of $\mathbf{D}_p(u)$ at x , Eq. (4.2) can be written as:

$$\mathbf{D}_m(x) = A \mathbf{D}_p(x), \quad (4.3)$$

where the operator is:

$$A = \exp\left(\frac{\sigma^2}{2} \frac{d^{(2)}}{dx^2}\right) = \sum_{n=0}^{\infty} \frac{1}{n!} \left(\frac{\sigma^2}{2}\right)^n \frac{d^{(2n)}}{dx^{2n}}.$$

The inverse operator A^{-1} , defined by Ulmer and Kaissl^[91], leads us to formulate the deconvolution from \mathbf{D}_m to \mathbf{D}_p as:

$$\mathbf{D}_p(x) = A^{-1} \mathbf{D}_m(x) = \exp\left(-\frac{\sigma^2}{2} \frac{d^{(2)}}{dx^2}\right) \mathbf{D}_m(x). \quad (4.4)$$

Ulmer and Kaissl also derived an expression of the inverse PDF or deconvolution kernel $p^{-1}(\sigma, x)$, which allows the deconvolution to be calculated in the following integral form:

$$\mathbf{D}_p(x) = \int p^{-1}(\sigma, x-u) \mathbf{D}_m(u) du. \quad (4.5)$$

In Eq. (4.6), the representation of $p^{-1}(\sigma, x)$ can be expressed in the expansion of Hermite polynomials H_{2n} :

$$p^{-1}(\sigma, x) = \sum_{n=0}^{\infty} c_n H_{2n}\left(\frac{x}{\sqrt{2}\sigma}\right) p(\sigma, x), \quad (4.6)$$

with the coefficients $c_n = (-1)^n / (2^n n!)$.

IV.2.2. 1D beam profile deconvolution

The Fourier transform of Eq. (4.4) can be expressed in the following form:

$$F\{A^{-1}\mathbf{D}_m(u)\} = \sum_{n=0}^{\infty} \frac{1}{n!} \left(\frac{\sigma^2}{2} \right)^n w^{2n} F\{\mathbf{D}_m(u)\} \quad (4.7)$$

where w is the angular frequency. Following Ulmer and Kaissl ^[91], the deconvolution kernel $p^{-1}(\sigma, x)$ can also be written as:

$$p^{-1}(\sigma, x) = A^{-2} p(\sigma, x), \quad (4.8)$$

where $A^{-2} = \exp\left(-\sigma^2 \frac{d^{(2)}}{dx^2}\right) = \sum_{n=0}^{\infty} \frac{1}{n!} (-1)^n \sigma^{2n} \frac{d^{(2n)}}{dx^{2n}}$. Then, the Fourier transform of the deconvolution kernel $p^{-1}(\sigma, x)$ is:

$$F\{p^{-1}(\sigma, x)\} = \sum_{n=0}^{\infty} \frac{1}{n!} \sigma^{2n} w^{2n} \exp(-w^2 \sigma^2 / 2). \quad (4.9)$$

Using the Fourier transform to Eq. (4.5), we obtain:

$$F\{\mathbf{D}_p(x)\} = \sum_{n=0}^{\infty} \frac{1}{n!} \sigma^{2n} w^{2n} \exp(-w^2 \sigma^2 / 2) F\{\mathbf{D}_m(u)\}. \quad (4.10)$$

By comparing Eq. (4.7) with Eq. (4.10), a filter $H(w)$ can be defined as:

$$F\{p^{-1}(\sigma, x-u)\mathbf{D}_m(u)du\} = H(w)F\{A^{-1}\mathbf{D}_m(u)\}, \quad (4.11)$$

where

$$H(w) = \frac{\sum_{n=0}^{\infty} \frac{1}{n!} \sigma^{2n} w^{2n}}{\sum_{n=0}^{\infty} \frac{1}{n!} \left(\frac{\sigma^2}{2} \right)^n w^{2n}} \exp(-w^2 \sigma^2 / 2). \quad (4.12)$$

Note that $H(w) = 1$ corresponds to an all-pass filter through the summation of polynomials to infinity. Due to the small contributions of the higher order terms, we only consider the first four terms in the summation of Eq. (4.12). In this context, $H(w)$ is a filter that suppresses part of the high-frequency components.

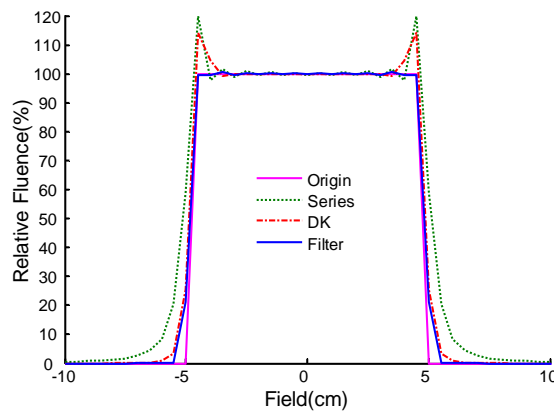


Figure 4.1. The profile for a static target and the corresponding convolved-back profiles for deconvolution profiles. “Origin”—nominal static profile, “Series”—profile obtained with the series expansion method in Eq. (4.4), “DK”—profile obtained with the deconvolution kernel method, “Filter”—profile obtained with the proposed method. A Gaussian distribution of geometrical uncertainty with standard deviation $\sigma = 3$ mm is assumed.

The convolved-back profiles for the deconvolution using the series expansion method in Eq. (4.4) and the deconvolution kernel method in Eq. (4.5) are shown in Fig. 4.1, where the series expansion method is labeled as “Series” and the deconvolution kernel method is labeled as “DK”. We found that there are large deviations around the edges of fields between the convolved-back profiles and the static profile. However, the convolved-back profiles using the deconvolution kernel method are better than the results of the series expansion method, because the convolved-back profiles from the deconvolution kernel method are equal to the product of the results from the series expansion method and a low-pass filter. The frequency response of this low-pass filter is related to the truncation of the terms considered in Eq. (4.12). In order to further reduce the differences between the convolved-back profiles and the static profile, we filtered the results of the series expansion method with a classical filter whose frequency response could be more easily adjusted. To this end, we used a Butterworth filter with a frequency response of:

$$H(w) = \frac{1}{\sqrt{1 + (w/w_0)^{2m}}}, \quad (4.13)$$

where w_0 is the cutoff frequency and m is the number of reactive elements (poles) in the filter.

In the series expansion (Eq. (4.8)) we just took the first four terms of the series—the low frequency components gathering most of the signals’ intensity. By adjusting the filter parameters, we could easily suppress the signals’ high-frequency components according to different PDF, and reduce differences between the static profile and the convolved-back profiles on the field edges, as shown in Fig. 4.1, where the proposed method is labeled as “Filter”.

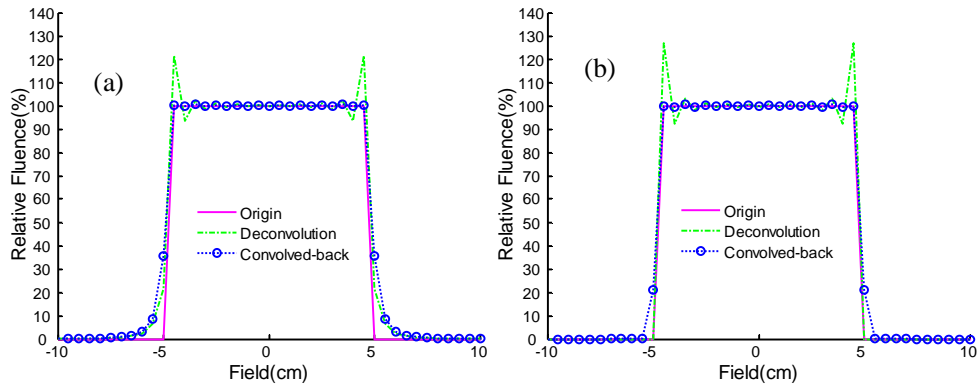


Figure 4.2. A rectangle-field profile: “Origin”—nominal profile for a static target, “Deconvolution”—the deconvolution profile, “Convolved-back”—the convolved-back profile. (a) The proposed method; (b) setting the value of the deconvolution profiles outside the field to zero.

To reduce the dose delivered outside the field, a further approximation was made by setting the values of deconvolution profiles outside the field to zero (Fig. 4.2). This may reduce the intensities of convolved-back profiles on the field edges, as no dose is received outside the field. Thus, it is necessary to readjust the filter parameters in order to compensate for these lower intensities by the increased dose received on the field edges, as shown in Fig. 4.2(b). The parameters of the filter were set by systematically exploring the parameter space for the minimal difference between the convolved-back profile and the

nominal static profile.

IV.2.3. Extension to 2D fluence deconvolution

The 2D extension can be derived following the same approach. For a 2D Gaussian PDF:

$$p(\sigma_x, \sigma_y, x, y) = \frac{1}{2\pi\sigma_x\sigma_y} \exp\left(-\frac{x^2}{2\sigma_x^2} - \frac{y^2}{2\sigma_y^2}\right), \quad (4.14)$$

where the variables x and y are independent and separable; therefore, the extension of the 1D operator equation (4.3) to 2D is straightforward:

$$A = \exp\left(-\left(\frac{\sigma_x^2}{2}\right)\frac{d^2}{dx^2}\right) \exp\left(-\left(\frac{\sigma_y^2}{2}\right)\frac{d^2}{dy^2}\right). \quad (4.15)$$

The deconvolution procedure is similar; the corresponding Fourier transform in each dimension is written as:

$$F\{\mathbf{D}_p(x, y)\} = \sum_{n=0}^{\infty} \frac{1}{n!} \left(\frac{\sigma_x^2}{2}\right)^n u^{2n} \sum_{k=0}^{\infty} \frac{1}{k!} \left(\frac{\sigma_y^2}{2}\right)^k v^{2k} F\{\mathbf{D}_m(x, y)\}, \quad (4.16)$$

where u and v are the angular frequency. To reduce fluctuations within the field, the deconvolution outputs were filtered by a 2D Butterworth filter. We wrote the 2D Butterworth filter in the frequency domain as follows:

$$H(u, v) = \sqrt{\frac{1}{1 + (u/u_0)^{2m}}} \sqrt{\frac{1}{1 + (v/v_0)^{2k}}}, \quad (4.17)$$

where u_0 and v_0 are the cutoff frequencies, and m and k are the number of reactive elements (poles) of the filter in each dimension.

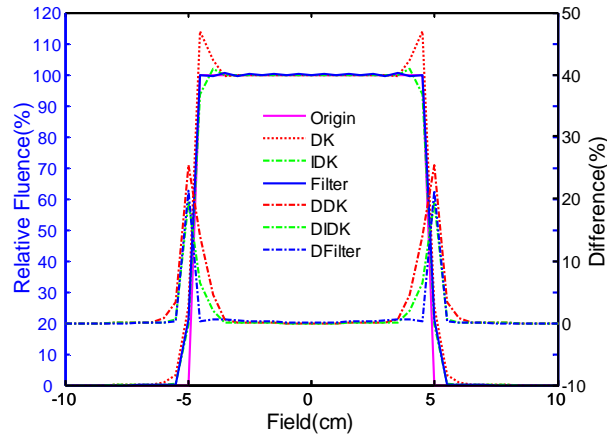


Figure 4.3. The profiles for a static target and the corresponding convolved-back profiles. “Origin”—nominal static profile, “DK”—the deconvolution kernel method, “IDK”—the improved deconvolution kernel method, “Filter”—the proposed method. The dash-and-dot lines present the corresponding absolute difference with static profile and symbol prefix “D” for each method name. Gaussian distribution of geometrical uncertainty with standard deviation $\sigma = 3$ mm is assumed.

In this work, the proposed method was compared to the improved deconvolution kernel method given by Fan and Nath ^[17]. That method achieves the deconvolution kernel by taking the first four Hermite polynomials in Eq. (4.6) in the Cartesian coordinate system and then effectuating an approximate conversion to the polar coordination system through a triangular transformation. Fan's method suppresses the fluence around the field edges. If the suppression is too large, however, it may lead to lower intensity in the convolved-back profiles than in the static fluence map around the field edges, as shown in Fig. 4.3.

IV.3. Results

We simulated two kinds of intensity-modulated 2D fluence maps, as shown in Fig. 4.4(a) and Fig. 4.5(a), and tested our method on a prostate case (Fig. 4.7(a)). The patient PDF is the PDF modeling random geometric uncertainties of the patient in 3D space. In this work, the PDF $p(\sigma_x, \sigma_y, x, y)$ is a projection of the patient PDF to a plane orthogonal to the beam direction. Standard deviations of $\sigma_x = 4\text{mm}$ and $\sigma_y = 3\text{mm}$ were chosen for 2D PDF in x - and y - directions, respectively (in Eq. (4.14)). These standard deviations are the typical random geometrical uncertainty in practice for various sites ^[95, 96]. To obtain a high cutoff frequency and a low order for the Butterworth filter, as discussed above, we used the parameters $u_0 = 0.9375$, $n = 1$ and $v_0 = 1.25$, $k = 1$ in Eq. (4.17), respectively, for these two standard deviations. Note that the values of cut-off frequencies u_0 and v_0 were selected by an exhaustive search in order to minimize the difference between the convolved-back fluence map and the static fluence map. The same filter parameters were employed for different fluence maps with equal standard deviations. Three different methods ("DK"--the deconvolution kernel method ^[91], "IDK"--the improved deconvolution kernel method ^[17], and "Filter"--the proposed method) were compared, both for the two kinds of fluence maps and the prostate case.

We calculated dose distribution as follows:

- (i) The deconvolution fluence maps were obtained via the deconvolution methods mentioned above;
- (ii) We sequenced the deconvolution fluence map into a series of deliverable leaf sequences, and then reconstructed these deliverable leaf sequences into a deliverable fluence map;
- (iii) The convolved-back fluence map was then calculated by convolving the deliverable fluence map and Eq. (4.14);
- (iv) Finally, the convolved-back fluence map was inserted into the software CERR ^[70], in order to calculate the dose distribution.

IV.3.1. Results for a regular 2D fluence map

Figure 4.4 shows the application of the 2D deconvolution algorithms to the regular 2D fluence map. Ideally, the convolved-back fluence maps (Fig. 4.4(b), (c) and (d)) should be the same as the nominal static one (Fig. 4.4(a)). The differences between the convolved-back fluence maps and the static fluence map lie mainly around the field edges

due to the nominal static fluence map's steep fall-off. A detailed comparison of convolved-back fluence maps is also shown in a profile plot in Fig. 4.4(e) for these three methods.

As shown in Fig. 4.4, the convolved-back fluence map based on the proposed method exhibits fewer hot or cold spots than the results with the other two methods. Much improved agreement on fluence distributions around the field edges was achieved. The difference between the nominal static fluence map and the convolved-back fluence map was measured by the quadratic sum defined as:

$$d = \sum \sum (\mathbf{F}_{ori} - \mathbf{F}_{con})^2, \quad (4.18)$$

where \mathbf{F}_{ori} is the nominal static fluence distributions and \mathbf{F}_{con} is the convolved-back fluence distributions. For this regular 2D fluence map, the quadratic sums of the difference between the nominal static map and the convolved-back fluence maps in the field were 0.7070, 0.4680, and 0.0577, for the deconvolution kernel method, the improved deconvolution kernel method, and the proposed method, respectively.

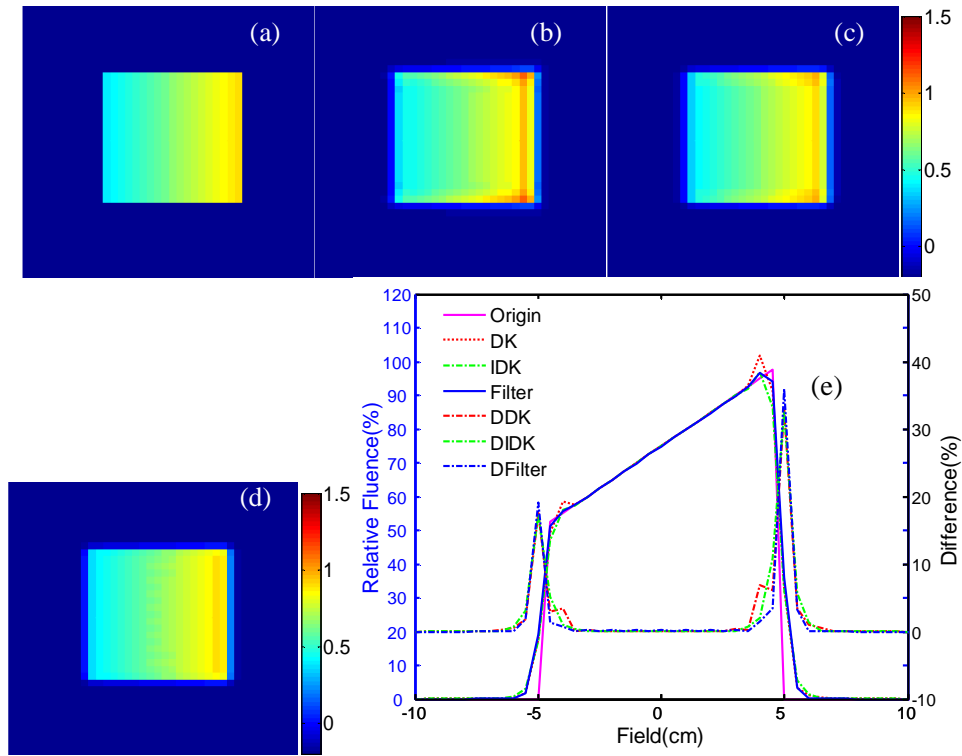


Figure 4.4. The fluence map for a static target and the corresponding convolved-back fluence maps from the deconvolution profiles. (a) The nominal static fluence map; the convolved-back fluence maps for: (b) the deconvolution kernel method, (c) the improved deconvolution kernel method, and (d) the proposed method; (e) the lateral profiles for these fluence maps.

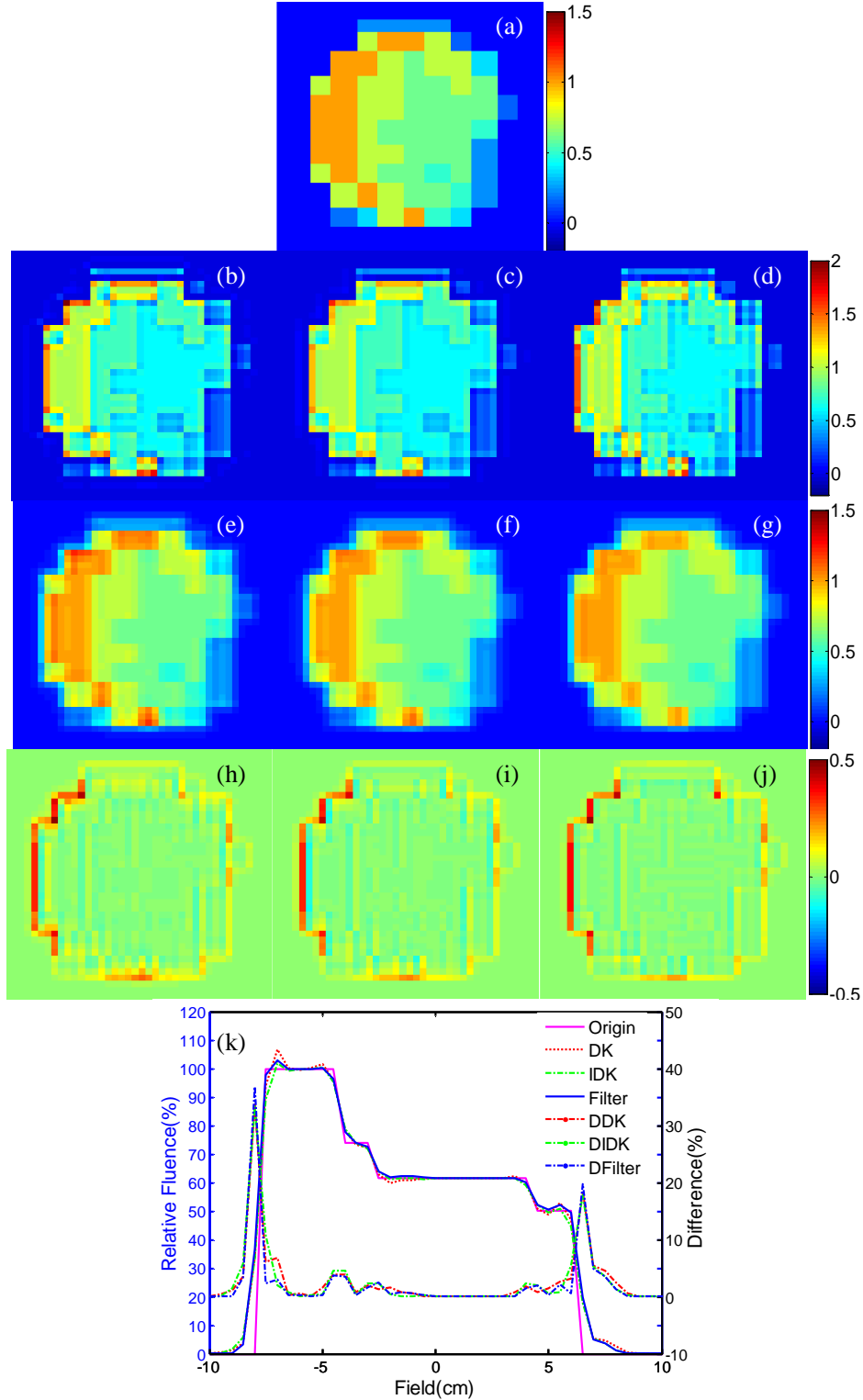


Figure 4.5. The 2D deconvolution algorithms to the fluence map of a real IMRT field. (a) the nominal static fluence map; (b)-(d) the deconvolution fluence maps from corresponding methods (the deconvolution kernel method, the improved deconvolution kernel method, and the proposed method, respectively); (e)-(g) the corresponding convolved-back fluence maps, respectively; (h)-(j) the corresponding difference between the convolved-back fluence maps and the nominal static fluence map respectively; (k) the lateral profiles for convolved-back fluence maps.

IV.3.2. Results for an IMRT field

Figure 4.5 shows the application of the 2D deconvolution algorithms to the fluence map of an IMRT field. Compared to the nominal static fluence map (Fig. 4.5(a)), the deconvolution fluence maps (Fig. 4.5(b)-(d)) displayed a higher intensity on the field edges, especially in the lateral direction (x -direction), due to a larger random geometrical uncertainty and the steeper intensity edges. The convolved-back fluence maps and the corresponding differences to the nominal static fluence map are depicted in Fig. 4.5(e)-(g) and 4.5(h)-(j). A profile comparison of the convolved-back fluence maps is shown in Fig. 4.5(k).

The deconvolution fluence map (Fig. 4.5(d)) computed using the proposed method is less smooth than those calculated using the other two methods (Fig. 4.5(b)-(c)). However, the convolved-back fluence map (Fig. 4.5(g)) is closer to the static fluence map than the two corresponding convolved-back fluence maps (Fig. 4.5(e)-(f)), especially on the field edges, as shown in Fig. 4.5(h)-(j). For the deconvolution kernel method, the improved deconvolution kernel method, and the proposed method, the quadratic sums of the difference between the nominal static fluence map and the convolved-back fluence maps were 1.2014, 1.0460, and 0.6884, respectively, in the field. As depicted in Fig. 5, the proposed method achieved superior agreement on fluence distributions around the field edges.

We input the convolved-back fluence maps (Fig. 4.5) into the software CERR [70] in order to calculate the dose distribution for a 6MV photon beam at 10cm depth of a water phantom, and the corresponding results are shown in Fig. 4.6, where Fig. 4.6(a)-(d) displays the dose distribution from the static fluence map in Fig. 4.5(a) and from the convolved-back fluence maps in Fig. 4.5(e)-(g). We used the gamma evaluation method [97] in order to analyze the dose distribution calculated from convolved-back fluence maps. The gamma analysis was applied with a dose difference of 3% and a distance to agreement of 3mm. The 2D gamma analysis is shown in Fig. 4.6(e)-(g), and the dose profiles are shown in Fig. 4.6(h). We observed that the dose distributions calculated using the proposed method were closer to the dose distribution calculated from the nominal static fluence map than those calculated using the other two methods, especially around the field edges.

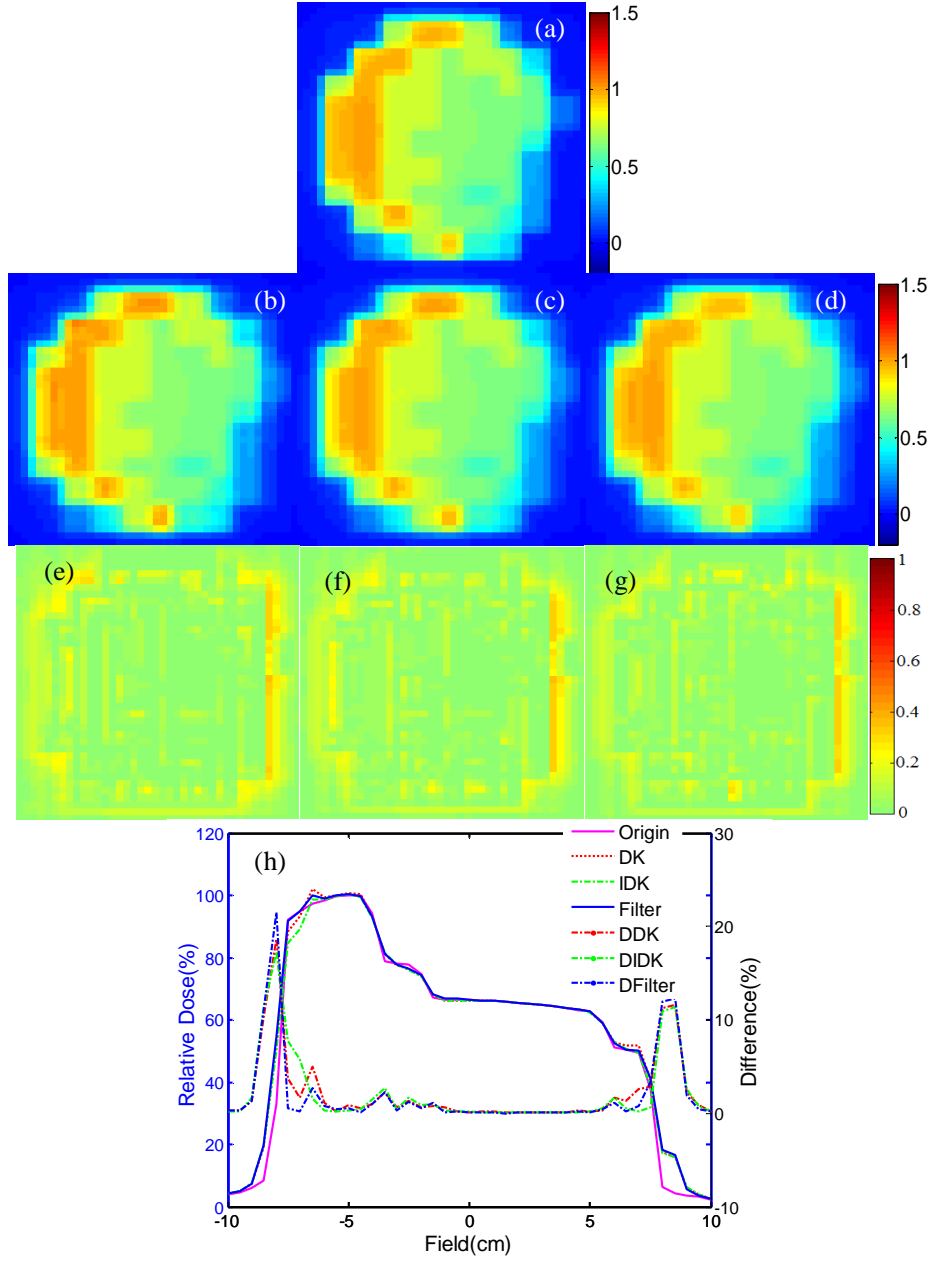


Figure 4.6. Dose distribution for a real IMRT field: (a) from the nominal static fluence map; (b)-(d) from the convolved-back fluence maps in Fig. 4.6(e)-(g) respectively; (e)-(g) the gamma analysis for the corresponding difference between Fig. 4.6(a) and Fig. 4.6(b)-(d), respectively; (h) the lateral profiles for dose distribution.

IV.3.3. Results for a prostate case

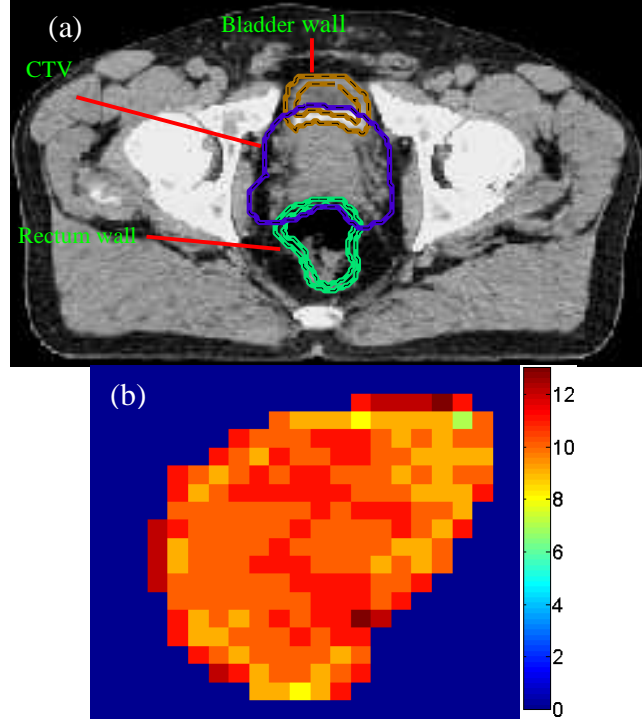


Figure 4.7. CT scans and fluence map. (a) The position of CTV and normal tissues in one CT slice of a patient; (b) the nominal static fluence map used in this case.

Our method was tested on a prostate case with one 6MV photon beam extracted from a 5-beam IMRT plan. The gantry angle was 90° , and the beam's source-axis distance was 100cm. Figure 4.7 displays one slice of the CT scans, along with the static fluence map we used. The sensitive normal tissues considered here were the rectum wall and bladder wall (Fig. 4.7(a)). We calculated dose distributions for both the nominal static fluence map and the fluence maps obtained using the margin expansion and deconvolution methods. The margin expansion method enlarged the static fluence map by 5mm in each direction. Fig. 4.8 shows the 2D dose distributions at 17cm under the surface in the beam direction for those fluence maps. We observed that high doses were partially delivered to the region outside the CTV when this approach was used (Fig. 4.8(b)). When the deconvolution kernel method was applied, the delivered dose was high around the target edges (Fig. 4.8(c)). The improved deconvolution kernel method (Fig. 4.8(d)) and the proposed method (Fig. 4.8(e)) showed a better dose modulation than the margin expansion method. Dose volume histograms (DVH) for the 2D dose distributions (Fig. 4.8) are displayed in Fig. 4.9. For the sensitive normal tissues, the proposed method offered better control for high doses delivered to normal tissue than the other three methods (Fig. 4.9(a) and (b)). For the CTV, the dose distributions obtained using the proposed method were closer to the nominal static fluence map than those resulting from the other three methods (Fig. 4.9(c)). We also analyzed the 2D dose distributions for the deconvolution method with gamma evaluation.

A significant reduction of the difference in dose distributions was achieved with the proposed method, as shown in Fig. 4.10.

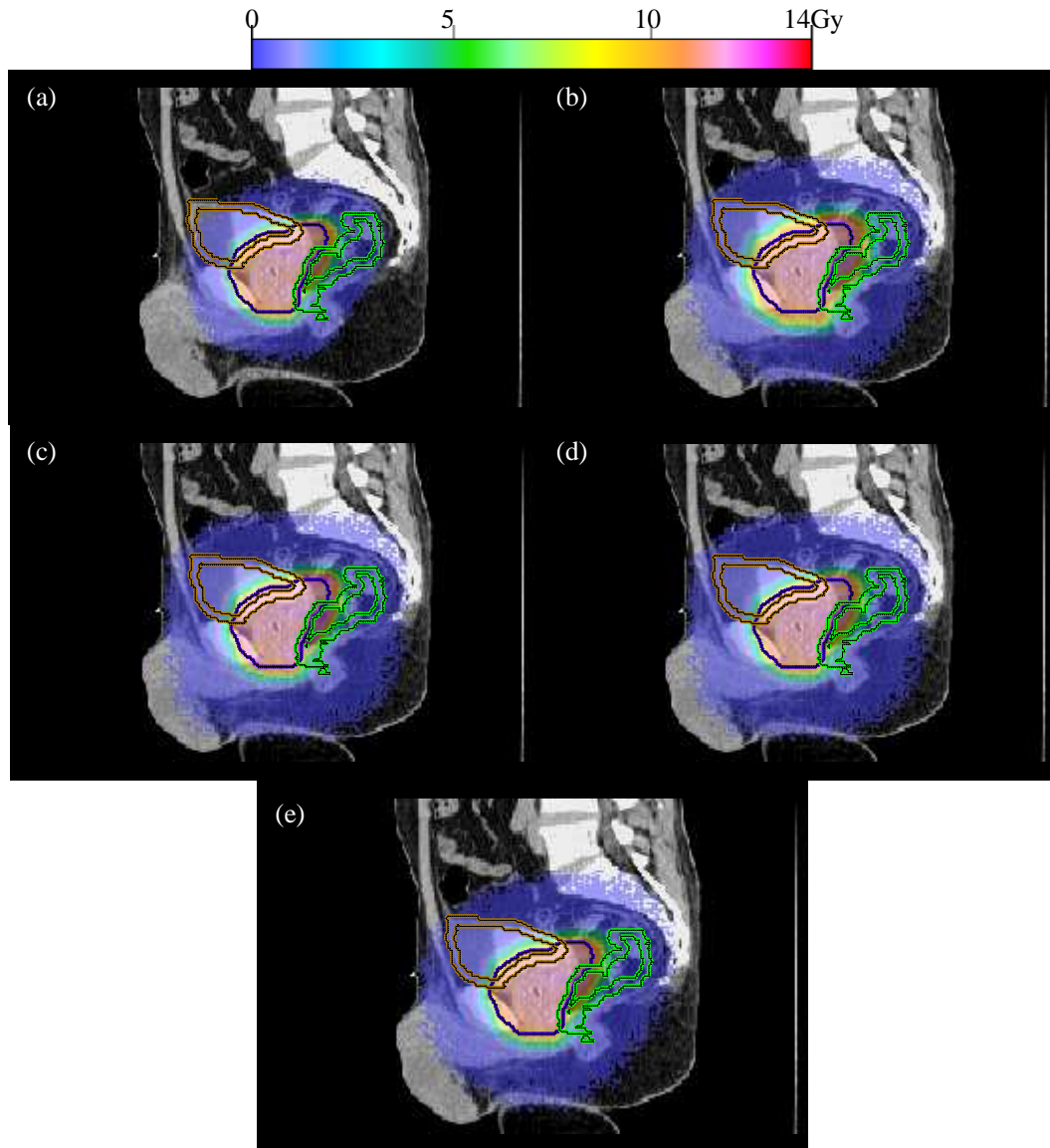


Figure 4.8. Dose distributions in the prostate case using: (a) the nominal static fluence map; and fluence maps resulting from (b) margin expansion method; (c) the deconvolution kernel method; (d) the improved deconvolution kernel method; (e) the proposed method.

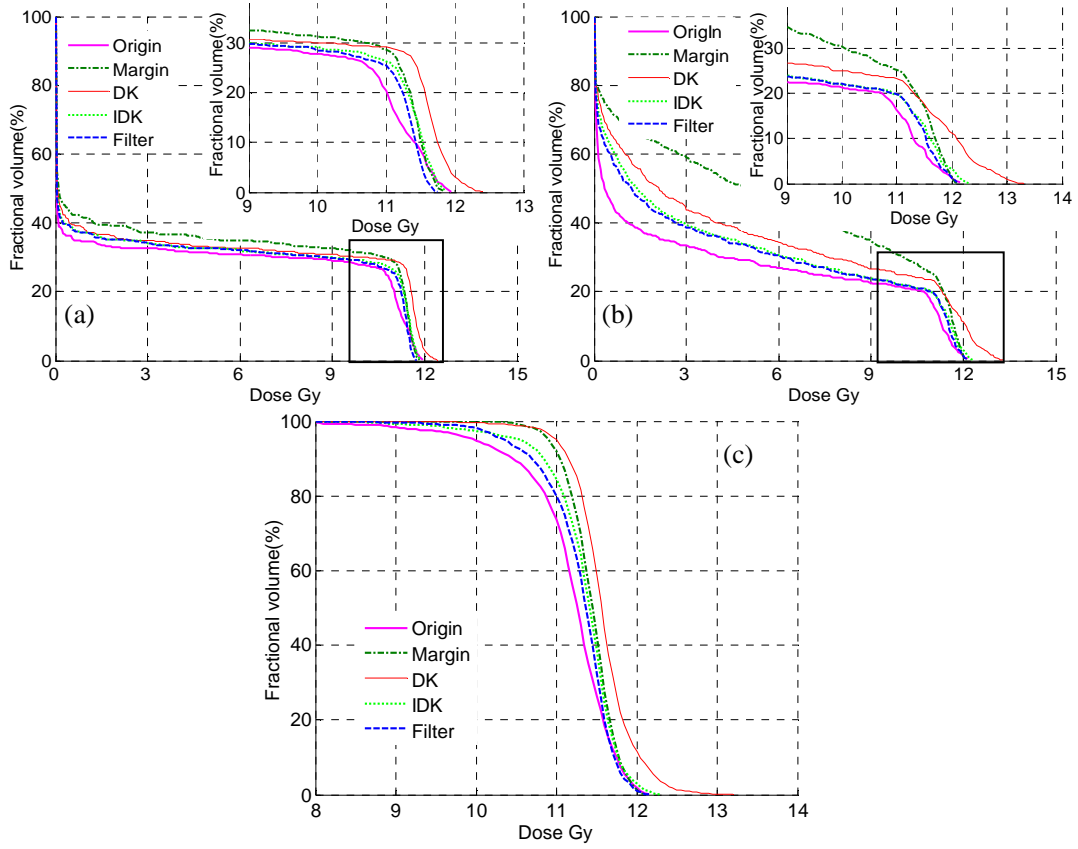


Figure 4.9. Dose-volume histograms of prostate case for the different methods with respect to the nominal static fluence map with: (a) bladder wall; (b) rectum wall; (c) CTV. “Margin” — margin expansion method.

We used the method proposed by Engel^[98] in order to sequence the fluence map into a deliverable leaf sequence. In this work, the pixel size in fluence map was $0.5 \times 0.5 \text{ cm}^2$, and the minimum monitor unit (MU) was 5MU for each segment. Thus, the resulting fluence maps were deliverable. With 10, 17, 14, and 16 segments, the fluence map monitor units equaled 85, 275, 175, and 210 for the margin expansion method, the deconvolution kernel method, the improved deconvolution kernel method, and the proposed method, respectively.

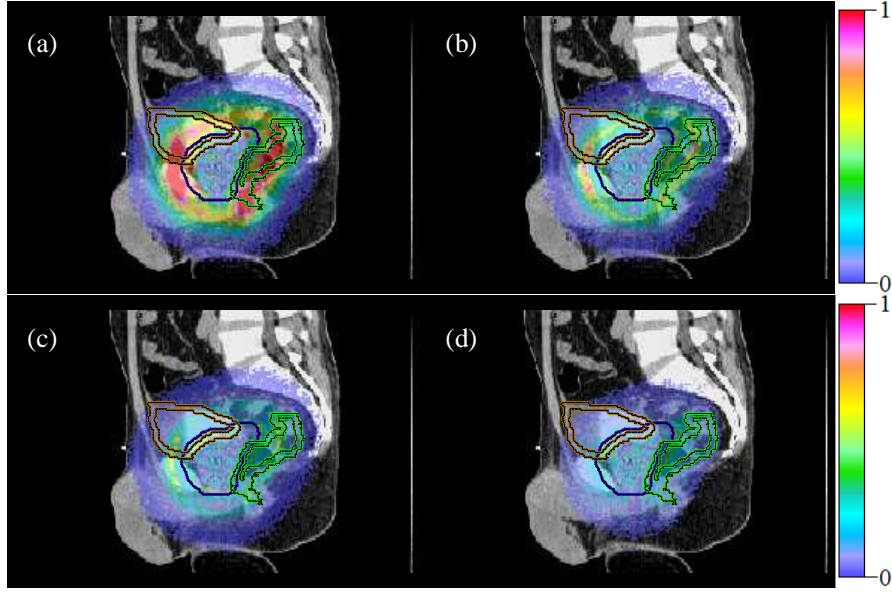


Figure 4.10. The gamma analysis for dose distributions in the prostate case using fluence maps resulting from: (a) margin expansion method; (b) the deconvolution kernel method; (c) the improved deconvolution kernel method; (d) the proposed method.

IV.4. Discussion and conclusion

This study's main assumption was that the organ motions of interest are random, with known probability distributions. Under this hypothesis, several works based on the deconvolution algorithm have been reported in the literature to derive a robust fluence map for IMRT. The advantages and disadvantages of this kind of approaches have been discussed by Fan and Nath^[17]. They showed that the deconvolution approach is more suitable for IMRT fluences with random geometrical uncertainties with a standard deviation within 2 and 4 mm, which means that the motion uncertainty is not very high. In this context, we have proposed a new deconvolution approach for robust fluence.

Our proposed solution combines series expansion and filtering techniques. Different optimal filters are available, the choice among which (Bessel, Chebyshev, elliptic, etc.) depends on several features, such as the behavior in pass-band, roll-off factor (this factor determines the extent of the transition between the pass-band and stop-band), phase response, and execution speed. Bessel filters, for instance, provide more sharp transition between pass-band and stop-band, but not the best phase response. The Butterworth filter selected here provides the flattest pass-band and a low roll-off factor. The main objective was, first, to show some of the benefits of the proposed combination. The use of other filters will be further explored in the future.

In addition, the proposed method offers the advantage that the whole formulation described operates in the frequency domain instead of the spatial domain. The

Butterworth filter can filter out the deviations around the field edges directly in the frequency domain. The parameters of the 2D Butterworth filter can be determined in each direction due to the independence and separability of arguments of PDF in the x- and y-directions. Their setting, varying with the PDF standard deviation, was designed to seek the minimum difference between the convolved-back fluence map and the static fluence map. It may be improved by replacing the exhaustive search with an optimization procedure.

Comparisons with two other deconvolution approaches ^[17, 91] show that the difference between the nominal static fluence map and the convolved-back fluence map, as measured by the quadratic sum, is significantly reduced with the proposed method. This advantage is somehow less pronounced when real IMRT data are processed.

The convolved-back fluence map presents the actual fluence delivered to a patient when motion is accounted for. Our goal was to cause the target to receive the same fluence during motion as the nominal static fluence map; namely, to reduce the deviations between the static fluence map and convolution-back fluence map. Usually, there are large gradients in the static fluence map. The proposed method increases TNMU complexity (i.e., total number of monitor units), as defined by Engel ^[98]. For the prostate case reported in this work, these complexities were 83.62MU, 268.05MU, 179.26MU, and 206.59MU for the margin expansion, deconvolution kernel, improved deconvolution kernel, and proposed method, respectively.

To conclude, this new approach using both series expansion and a Butterworth filter was tested on two 2D fluence maps and a prostate case. It was shown that the Butterworth filter better suppressed high-frequency oscillations and reduced hot and cold spots on convolved-back fluence profiles. In the flat area of the fluence map, our method's accuracy was similar to that of the deconvolution kernel method. Near the edge of the fluence map, a clear advantage was observed. This approach can be easily implemented in the clinical setting and thus improve dose homogeneity.

The methods proposed in these three chapters have a common goal: to precisely optimize the plan by taking into both the physical (dose calculation), biological (NTCP constraints) and anatomical (correction of uncertainties) components. However, during the treatment delivery, some systematical anatomical variation may occur, e.g. related to the tumor or parotids shrinkage in head & neck RT. In that case, some replannings may be considered in an adaptive RT workflow. This is the purpose of the next chapter.

Chapter V. Assessment of the optimal time and number of replannings in head-and-neck cancer adaptive radiotherapy to spare the parotid glands

V.1. Introduction

Intensity-modulated radiotherapy plays a critical role in the current management of patients especially, with head-and-neck cancer. IMRT maximizes tumor coverage and sparing of organs at risk (OAR) and thus leads to a potential increase in the therapeutic index^[19]. However, the location, shape, and size of target and normal tissue are altered significantly during a 6-7 weeks course of treatment due to daily external daily setup variations, internal anatomical variations (physiological or in response to the irradiation) and weight loss^[19-22]. These changes include shrinkage of the tumor and normal tissue volumes, as well as positional shift of some of these structures^[22, 23, 99-101]. These anatomical changes can lead to overdose the healthy tissue and therefore to unexpected high normal tissue complications^[19, 23, 24]. The highly conformal IMRT plan generated on the initial planning CT may no longer be optimal during the treatment. One of the main OAR is the parotid gland (PG), whose irradiation leads to xerostomia. Indeed, xerostomia, induces difficulties in swallowing and speaking, loss of taste, and dental caries, with therefore a direct impact on patient quality of life.

In order to account for daily setup errors and anatomical variations, an off-line scheme can be used to effectively correct and compensate for anatomical variations in head-and-neck adaptive radiotherapy, i.e. replanning(s) realized between treatment fractions to spare the PGs at least as they were spare at the planning. This scheme seems appropriate for this localization because the considered anatomical variations are quite regular during the course of treatment. Replanning is however a heavy task, which includes acquiring CT scans, delineating organs, and optimizing a plan. Due to the limited resources, replanning right before each fraction is not realistic and likely unnecessary for all the patients. Indeed, the optimal number and time of replannings need to be identified. This issue has been previously explored but remains unsolved^[102-105].

It must be pointed out that this work results from a collaboration between several actors: radiation oncologists (for the delineation), physicists (for the planning) and PhD students (registration). The contribution of the present work is mainly in the analysis of the replannings.

V.2. Purpose and general framework of the study

In the context of IMRT for locally advanced HNC, this study aimed to:

- estimate the difference between the planned dose and the actual delivered dose to the PGs without replanning (PG overdose);
- assess the benefits of different replanning strategies to spare the PGs, at least as they were spared at the planning.

In summary (Figure 5.1), 15 patients received radical IMRT (70Gy) for locally advanced head and neck cancer. After the initial planning CT, 5 weekly CTs were used to estimate the dose distributions delivered during the treatment according to two scenarios, corresponding either to the initial planning (without replanning to assess the PG overdose) or to weekly replanning(s) (to assess the benefit of ART). PGs dose were recalculated therefore at the fraction, from the weekly CT. A total of 31 possible replanning combinations were tested (Table 5.1.). For each scenario, the accumulated dose during the whole treatment was estimated in the PGs (Figure 5.2). For this purpose, the deformation fields from the weekly CT scans to the initial CT scan were firstly obtained by using the symmetric forces demons algorithm^[106]. Secondly, weekly replan dose distributions were warped by the resulting deformation fields. Thirdly, the PG cumulated doses were computed for all scenarios. At last, the optimal scenario was identified by comparing the PG dosimetric endpoints between the different replanning scenarios.

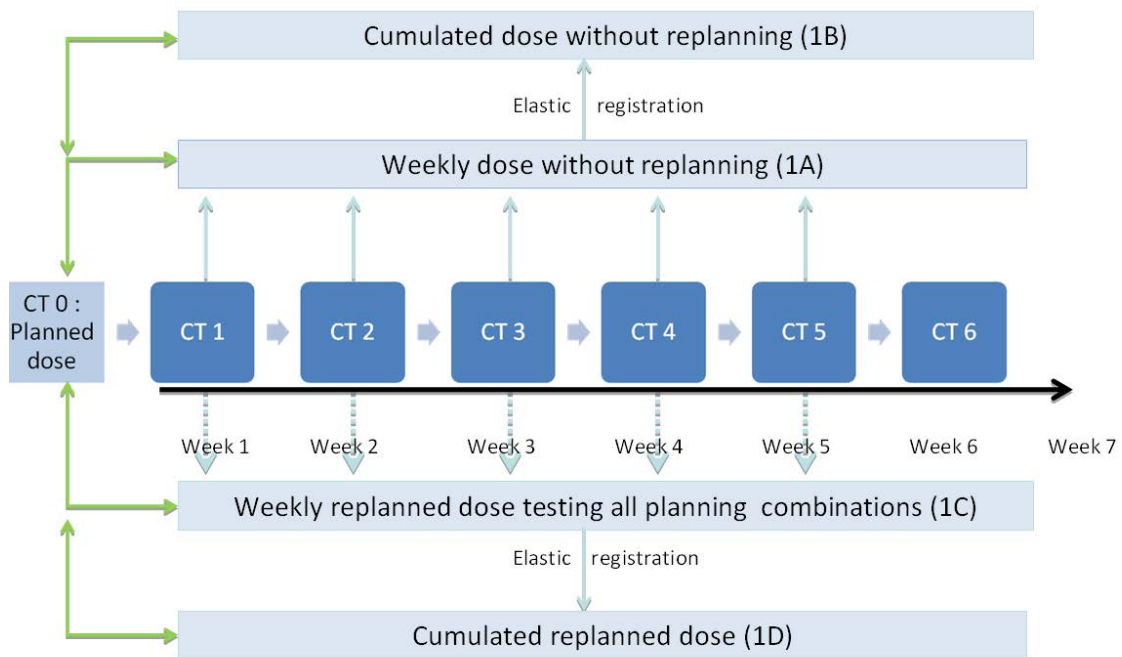


Figure 5.1. Study flow chart.

		Number and time of replannings				
		Once	Twice	Three times	Four times	Five times
Combinaisons	1	Replan 1	Replan 1 2	Replan 1 2 3	Replan 1 2 3 4	Replan 1 2 3 4 5
	2	Replan 2	Replan 1 3	Replan 1 2 4	Replan 1 2 3 5	
	3	Replan 3	Replan 1 4	Replan 1 2 5	Replan 1 2 4 5	
	4	Replan 4	Replan 1 5	Replan 1 3 4	Replan 1 3 4 5	
	5	Replan 5	Replan 2 3	Replan 1 3 5	Replan 2 3 4 5	
	6		Replan 2 4	Replan 1 4 5		
	7		Replan 2 5	Replan 2 3 4		
	8		Replan 3 4	Replan 2 3 5		
	9		Replan 3 5	Replan 2 4 5		
	10		Replan 4 5	Replan 3 4 5		

Table 5.1. The number and time of replannings used for each scenario (ex.: Replan 2,3,5 means : 3 replannings performed at weeks 2, 3 and 5).

V.3. Materials

The data of head-and-neck cancer patients were acquired in the Centre Eugène-Marquis hospital in Rennes, France.

V.3.1. Patients and tumors

The study enrolled a total of 15 patients with a mean age of 65 years (range: 50-87 years). All tumors were locally advanced (Stage III or IV, AJCC 7th ed). The mean PG volume was 25.3cc (Standard deviation (SD): 8.1cc), and the mean PG planned dose was 30.9Gy (SD: 7.9Gy).

V.3.2. Treatment and planning

All patients underwent IMRT using a total dose of 70Gy (2Gy/fraction/day, 35 fractions), with a simultaneous integrated boost technique and concomitant chemotherapy. Planning CTs (CT0) with intravenous contrast agents were acquired with 2mm slice thickness from the vertex to the carina. A thermoplastic head and shoulder mask with five fixation points was used. PET-CT and MRI co-registration was used for tumor delineation. Three target volumes were generated. Gross tumor volume (GTV) corresponded to the primary tumor along with involved lymph nodes. Clinical target volume 70Gy (CTV₇₀) was equal to GTV plus a 5mm 3D margin, which was adjusted to exclude air cavities and bone mass without evidence of tumor invasion. CTV₆₃ corresponded to the area at high-risk of microscopic spread, while CTV₅₆ corresponded to the prophylactic irradiation area. GTV, CTV₆₃, CTV₅₆, and all organs at risk were manually delineated on each CT slice. Adding a 5mm 3D margin around the CTVs generated the PTVs. PTV expansion was limited to 3mm from the skin surface in order to avoid the build-up region and to limit skin toxicity [15]. All IMRT plans were generated using Pinnacle V9.2. Seven Coplanar 6-MV

photon beams were employed with a step and shoot IMRT technique. The prescribed dose was 70Gy to PTV₇₀, 63Gy to PTV₆₃, and 56Gy to PTV₅₆. The collapsed cone convolution/superposition algorithm was used for dose calculation. The maximum dose within the PTV was 110% (D_{2%}). The minimum PTV volume covered by the 95% isodose line was 95%. Dose constraints were set according to the GORTEC guidelines: a mean dose (D_{mean}) <30Gy and a median dose <26Gy for contralateral PGs.

V.3.3. Weekly dose estimations in cases of replanning

During the treatment, each patient underwent six weekly CTs (CT1 to CT6) according to the same modalities as CT0, except for the intravenous contrast agents (not systematically used, particularly in case of cisplatin based chemotherapy). For each patient, the anatomical structures were manually segmented on each weekly CT by the same radiation oncologist. In case of complete response, initial macroscopically-involved areas were still included in the CTV₇₀, which was adjusted to exclude any air cavities and bone mass without evidence of initial tumor invasion.

Fraction doses were estimated by calculating the dose distribution on the 5 weekly CTs, using treatment parameters and isocenter from CT0. Weekly re-planned doses were calculated by generating a new IMRT plan on each weekly CT in accordance with the dose constraints described for the initial planning. PTV coverage did not differ between initial planning and weekly re-planned CT. The dose constraints for the organs at risk have respected the GORTEC recommendations at the initial planning and in all replanning.

V.4. Methods for dose accumulation

V.4.1. Registration algorithm

Image registration can be used to monitor disease progression or response to treatment. Registration aligns CT scans by applying a geometrical transformation to map voxels from one CT scan to the homologous voxels in the other CT scan. In this work, the registration was implemented in two steps, the rough registration and the fine registration. In the rough registration, the spatial mapping was calculated to align the moving CT scans (weekly CTs) with the fixed CT scan (planning CT). For this purpose, a rigid registration algorithm was used. In the fine registration, a deformation field was estimated to bring the processed moving CT scan into alignment with the fixed CT scan. For this purpose, the symmetric forces demons algorithm ^[106] was applied to get the deformable fields.

The symmetric forces demons algorithm ^[106] is one kind of non-linear image registration, which solves the computational complexity and memory requirements of the estimation of large deformations. Non-linear image registration aims at finding a well-behaved spatial transformation s that best aligns two given data sets **I** and **J**. At each

iteration, non-parametric methods need to find the displacement $\mathbf{s}(\mathbf{p})$ of each point \mathbf{p} in order to improve the following global energy function:

$$E(\mathbf{I}, \mathbf{J}, \mathbf{c}, \mathbf{s}) = \frac{1}{\sigma_i^2} \text{Sim}(\mathbf{I}, \mathbf{J}, \mathbf{c}) + \frac{1}{\sigma_x^2} \text{dist}(\mathbf{c}, \mathbf{s})^2 + \frac{1}{\sigma_T^2} \text{Reg}(\mathbf{s}), \quad (5.1)$$

where \mathbf{c} is the point corresponding to \mathbf{p} , $\text{Sim}(\mathbf{I}, \mathbf{J}, \mathbf{c}) = \|\mathbf{I} - \mathbf{J} \circ \mathbf{c}\|^2$ is a similarity criterion used to measure the resemblance of the aligned images, $\text{dist}(\mathbf{c}, \mathbf{s}) = \|\mathbf{c} - \mathbf{s}\|$, $\text{Reg}(\mathbf{s}) = \|\nabla \mathbf{s}\|^2$ is a regularization energy which estimates the likelihood of the transformation, σ_i accounts for the noise on the image intensity, σ_x accounts for a spatial uncertainty on the correspondences, and σ_T controls the amount of regularization we need. In order to guarantee the invertibility of the deformation and to have access to the true inverse transformation, the symmetric forces demons algorithm works completely in the log-domain, i.e. it uses a stationary velocity field \mathbf{v} , $\mathbf{s} = \exp(\mathbf{v})$. To be consistent with the log-domain representation and keep the simplicity of the demons algorithm, one part of the sub-energy function can simply be written as $\text{dist}(\mathbf{c}, \mathbf{s}) = \|\log(\mathbf{s}^{-1} \circ \mathbf{c})\|$ and $\text{Reg}(\mathbf{s}) = \|\nabla \log(\mathbf{s})\|^2$. At each iteration, the stationary velocity field \mathbf{v} is updated in the form $\mathbf{v} + \mathbf{u}$ by solving the optical flow function:

$$E_{\text{diffeo}}^{\text{corr}}(\mathbf{I}, \mathbf{J}, \mathbf{v}, \mathbf{u}) = \text{Sim}(\mathbf{I}, \mathbf{J}, \exp(\mathbf{v}) \circ \exp(\mathbf{u})) + \|\mathbf{u}\|^2. \quad (5.2)$$

In order to test the ability of the demons deformable registration algorithm we used to globally align the organs, the dice score values are calculated for the target and normal tissues. The positions of organs for all weekly CT scans are firstly transformed with corresponding deformation field to get the aligned weekly organs $\mathbf{A}_{\text{organ}}$. Then, the dice score is computed as

$$\text{Dice} = \frac{2|\mathbf{A}_{\text{organ}} \cap \mathbf{B}_{\text{organ}}|}{|\mathbf{A}_{\text{organ}}| + |\mathbf{B}_{\text{organ}}|} \quad (5.3)$$

where $\mathbf{B}_{\text{organ}}$ is the organs' position for the initial CT scans.

V.4.2. Aligning the dose distributions

As mentioned above, the spatial mapping of two CT scans was computed by the rigid registration method and the deformation field \mathbf{s} was obtained by using the symmetric forces demons algorithm^[106]. Hence, the alignment of weekly dose distributions with the initial dose distributions was also applied in two steps. At first, the spatial mapping was used to align the weekly dose distributions to the initial dose distributions roughly. Then, the processed weekly dose distributions were wrapped to the initial dose distributions by the deformation field. The deformation field is represented as a three dimension vector field, whose voxel is a vector with the same dimension as the input data. Each vector in the deformation field represents the transformation between a point \mathbf{p}_{in} in the input space and a point \mathbf{p}_{out} in the output space such that:

$$\mathbf{p}_{in} = \mathbf{p}_{out} + \mathbf{s}. \quad (5.4)$$

The warped weekly dose distributions were calculated by:

$$\mathbf{D}_{out}(\mathbf{p}_{out}) = \mathbf{D}_{in}(\mathbf{p}_{in} - \mathbf{s}) \quad (5.5)$$

where \mathbf{D}_{in} is the processed weekly dose distributions, and \mathbf{D}_{out} is the deformed weekly dose distributions. Typically the mapped positions did not correspond to an integer pixel positions in the input data. Linear interpolation method was used to compute values at non-integer positions.

V.4.3. Calculating the accumulated dose

After these two steps of registration, all the weekly dose distributions were aligned to the initial anatomy. For each scenario (cf. Table 5.1), the accumulated dose \mathbf{D}_{cum} was computed by:

$$\mathbf{D}_{cum} = \sum_{i=1}^N g_i \mathbf{D}_i, \quad (5.6)$$

where N is the number of plans used in this scenario, \mathbf{D}_i is the initial dose distributions or the aligned weekly dose distributions, and g_i is the duration of the corresponding dose distribution used in the treatment. Due to the time sequence, the current week cannot use the replan made after this week. The accumulated dose process is presented in Fig. 5.2.

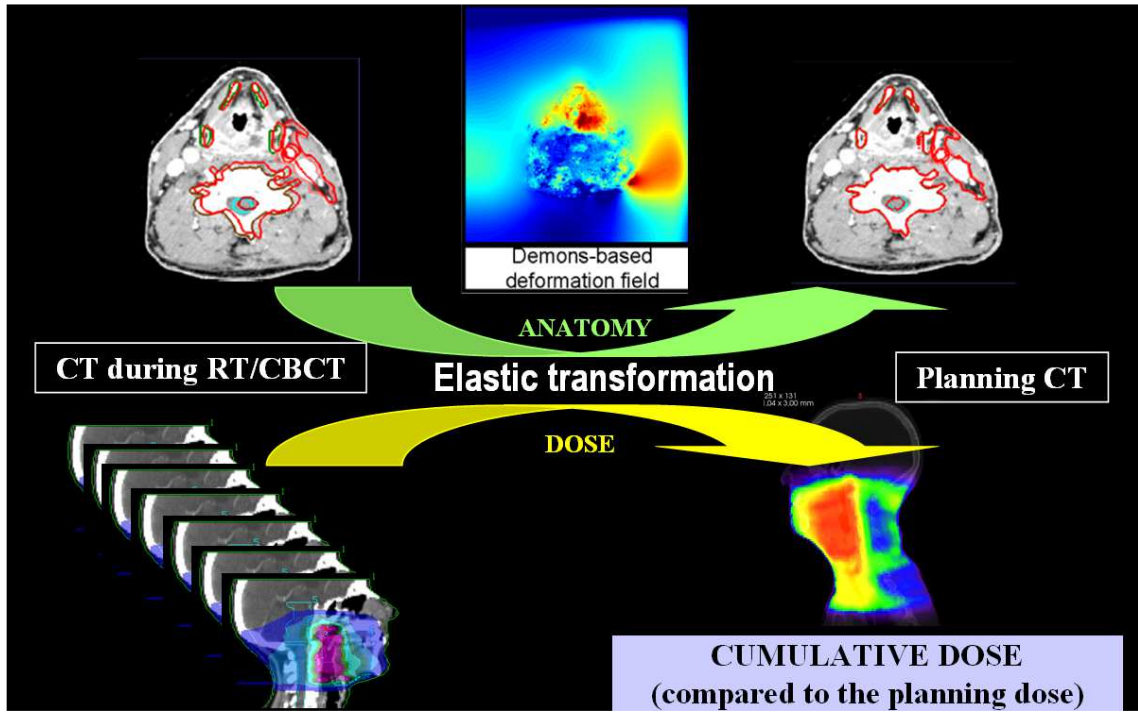


Figure 5.2. Cumulative dose workflow

V.4.4. Analyzing the cumulated dose and estimating the optimal scenario

The goal of radiation therapy is to deliver a prescribed radiation dose to the target volume while sparing surrounding functional organs and normal tissues, in particular the

parotid gland. The quality of one scenario was analyzed by the dosimetric indicators, such as DVH curves, mean dose, EUD and NTCP. The risk of xerostomia was estimated by using the LKB NTCP model ($n=1$, $m=0.4$, and $TD_{50}=39.9$), the complication being defined as a salivary flow ratio $<25\%$ of the pretreatment one

These indicators were calculated for the accumulated doses of all scenarios. The best scenario for a normal tissue was the one minimizing the dose it received, while for the target the goal was to minimize the difference between the prescribed dose and the delivered dose.

However, for one given patient, the best scenario for normal tissue and for target could be not the same one. Therefore, an objective function F was used to evaluate the overall result of each scenario as follows:

$$F = \frac{\sum_{i=1}^{N_{organ}} w_i f_i(\mathbf{D}_{cum})}{\sum_{i=1}^{N_{organ}} w_i} \quad (5.7)$$

where N_{organ} is the number of organs considered, f_i is the sub-function for the i th organ, such as mean dose, w_i is the weight which represents the importance of the i th organ. In this work, the mean dose of the accumulated dose \mathbf{D}_{cum} was used to evaluate the dose for each organ,

$$f_{mean}(\mathbf{D}) = \frac{1}{N_0} \sum_{i=1}^{N_0} \mathbf{D}_{cum,i} \quad (5.8)$$

where N_0 is the number of voxels in the anatomic structure of interest, $\mathbf{D}_{cum,i}$ is the dose in the i th voxel. For a normal tissue, the sub-function f_i indicates the dose delivered,

$$f_i(\mathbf{D}_{cum}) = f_{mean}(\mathbf{D}_{cum}) \quad (5.9)$$

For the target, the sub-function f_i reflects the absolute difference between the prescribed and the dose delivered,

$$f_i(\mathbf{D}_{cum}) = \text{abs}(f_{mean}(\mathbf{D}_{cum}) - D_{prescribed}) \quad (5.10)$$

where $D_{prescribed}$ is the prescribed dose to target. Let \mathbf{C} be the set of all scenario in Table 5.1. The optimal scenario is the scenario with the smallest objective function value

$$\min_{\beta \in \mathbf{C}} (F_{\beta}) \quad (5.11)$$

The mean dose (D_{mean}) was used to evaluate the dose delivered to organs. A non-parametric test (Wilcoxon test) was used for significance level comparison.

V.5. Results

V.5.1. Anatomical variations occurring during the 7 weeks of treatment

- From CT0 to CT6, the PG volumes decreased by a mean value of 28.3% (ranging from 0.0 to 63.4%), corresponding to an average decrease of 1.1cc/week (ranging from 0.0 to 2.2cc/week). The CTV₇₀ decreased by a mean value of 31% (ranging from 73% to -13%). Figure 3 shows the PGs and CTV₇₀ volume evolution in the course of treatment.

- The distance between the PGs and the CTV₇₀ (PG-CTV_{ds}) decreased by 4.3mm on average (ranging from 0 to 12mm) in 74% of the PGs, whereas it increased by 3.2mm on average (ranging from 1.1 to 6.3mm) in the other 26% of the PGs.

- The thickness of the neck at the level of the PG decreased by a mean value of 7.9mm (ranging from 0.1 to 26.6mm, SD 6.2mm) for 78% of the patients.

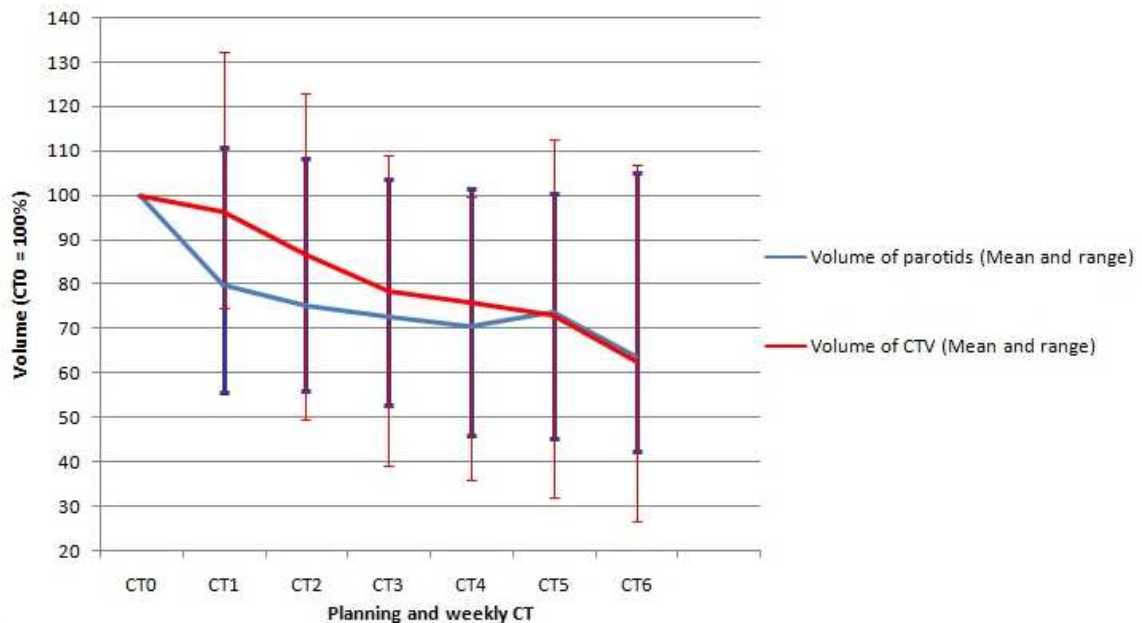


Figure 5.3. Reduction from CT0 (planning CT) to CT6 (CT at 6th week of treatment) of the PGs volumes and the CTV₇₀

V.5.2. Accumulated parotid gland dose distributions without replanning (compared to the planning)

By comparing the doses without replanning to the planned doses, the mean dose (D_{mean}) increased in 59% (N=16) of the PGs. Figure 5.4.a shows the D_{mean} difference for each PG of each patients. Ten out of 15 patients received a higher D_{mean} in at least one PG (6 patients in the 2 PGs). A dose increase was observed in 58% (N=7) of ipsilateral PGs and 60% (N=9) of contralateral PGs, which corresponded to a D_{mean} increase of an average of 3.7Gy (from 0.4 to 10.0Gy). Figure 5.4.b shows the corresponding difference in the estimated xerostomia risk. The average absolute increased risk of xerostomia was 3% (ranging from -16.7 to 23.9%) in all patients, and was 8.2% (ranging from 3.8 to 23.9%) among patients with an increased dose to PGs.

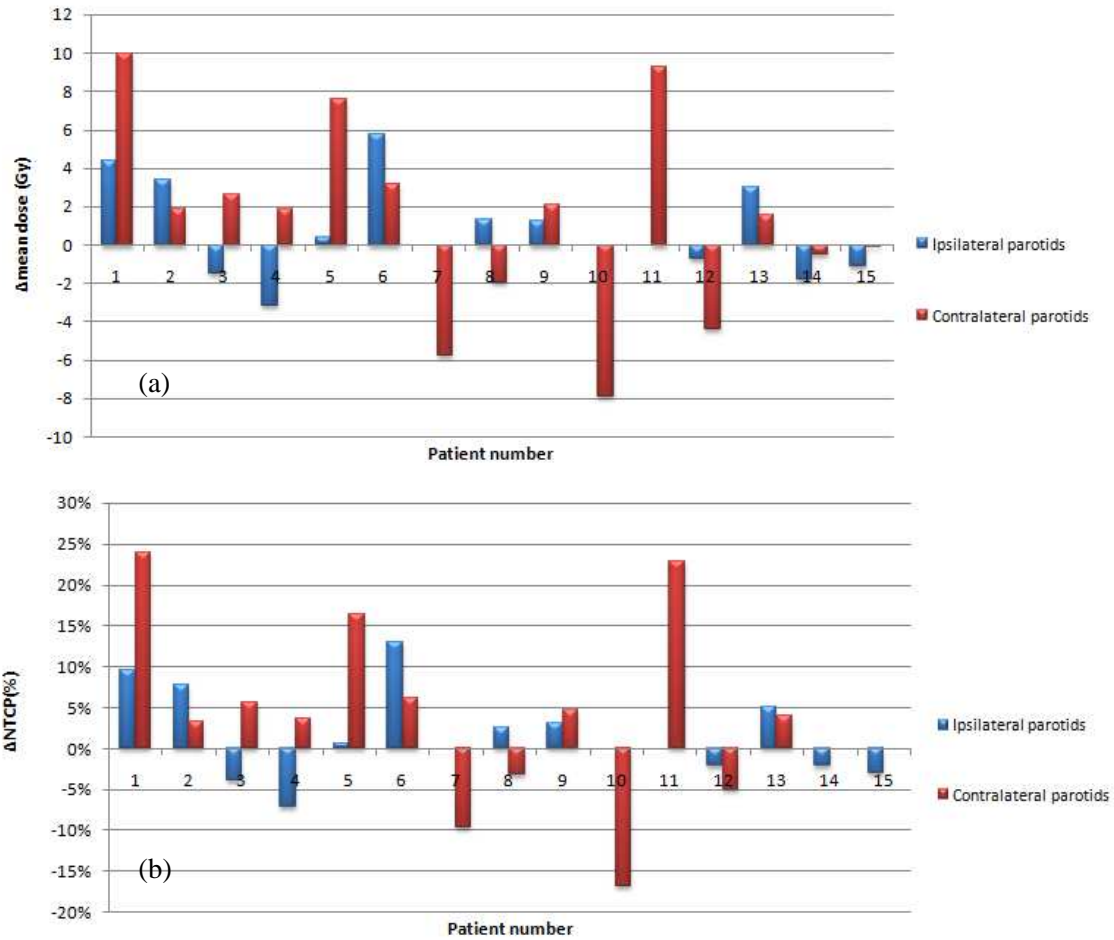


Figure 5.4. Difference between the actual delivered dose and the planning, for each patient and PG is presented: (a) mean dose (Gy) and (b) estimated xerostomia risk (%)

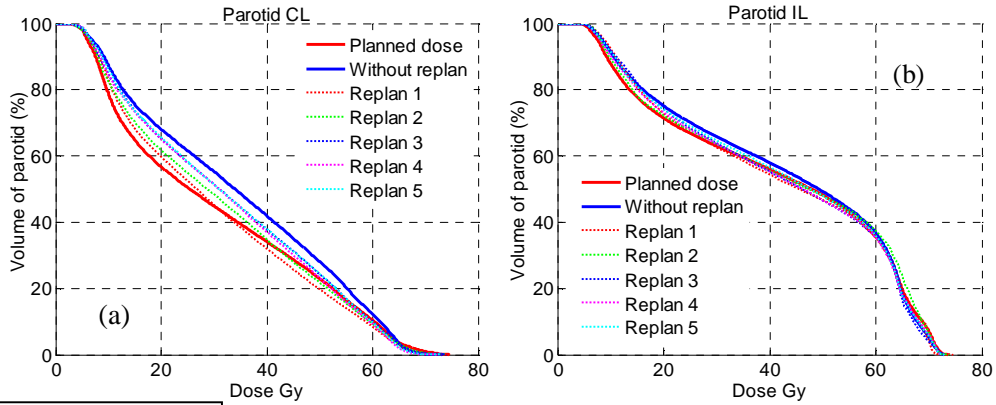
V.5.3. Accumulated parotid gland dose distributions corresponding to 31 scenarios of replannings

Thirty scenarios of different numbers and times of replannings have been tested, as shown in Table 5.1. Figure 5.5 displays the DVH curves of the accumulated dose for these 31 scenarios of all the patients, in the ipsilateral and contralateral PG. The DVH curves of the initial dose (Planned dose) and the accumulated dose of the scenario without replanning (Without replan) are also shown.

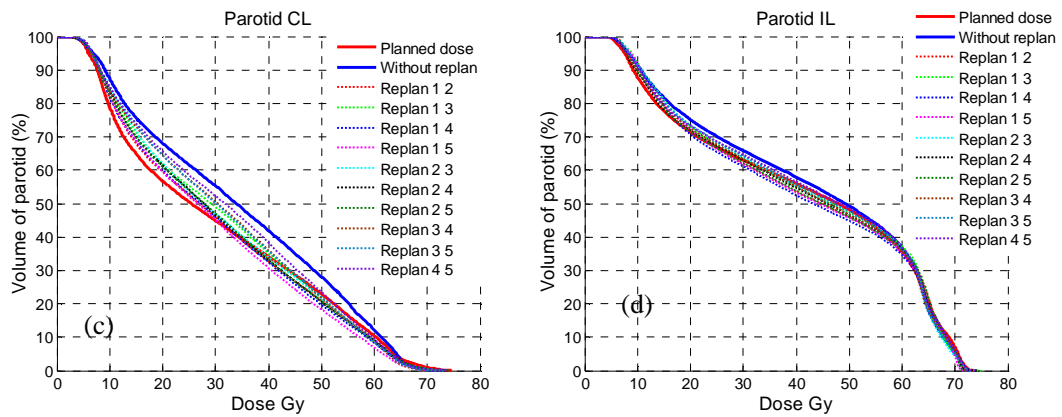
It can be observed that the replannings decreased more the doses delivered to the contralateral parotid gland than those to the ipsilateral parotid gland. Indeed, for the contralateral parotid gland, the distance to the target is larger. Thus, the replannings could spare it from high dose regions. For the ipsilateral parotid gland, however, part of or the entire gland was covered by the PTV. So, by keeping the same tumor coverage, it was

difficult to decrease the dose delivered to normal tissue.

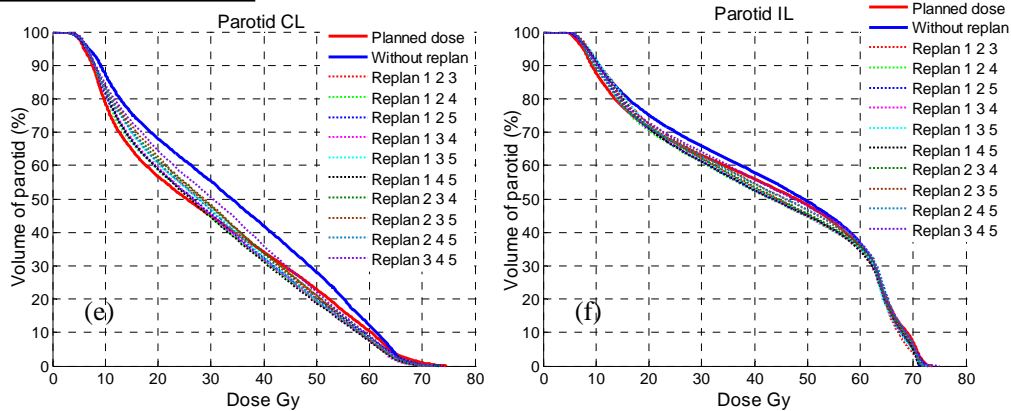
One replanning



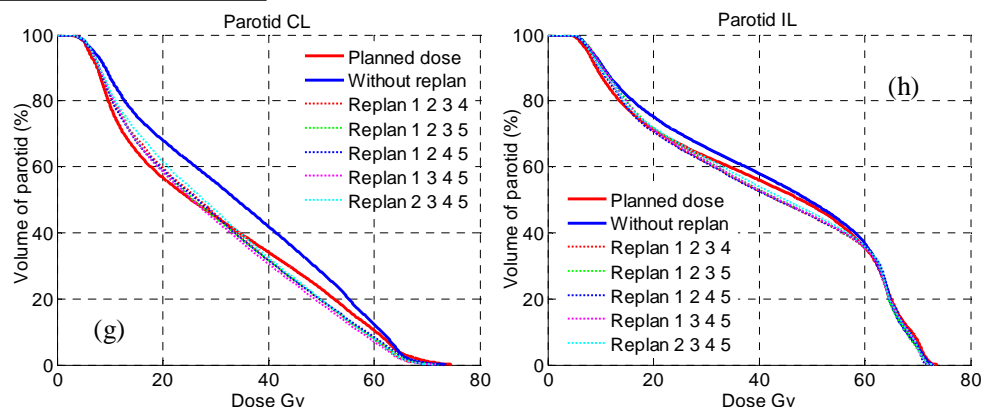
Two replannings



Three replannings



Four replannings



Five replannings

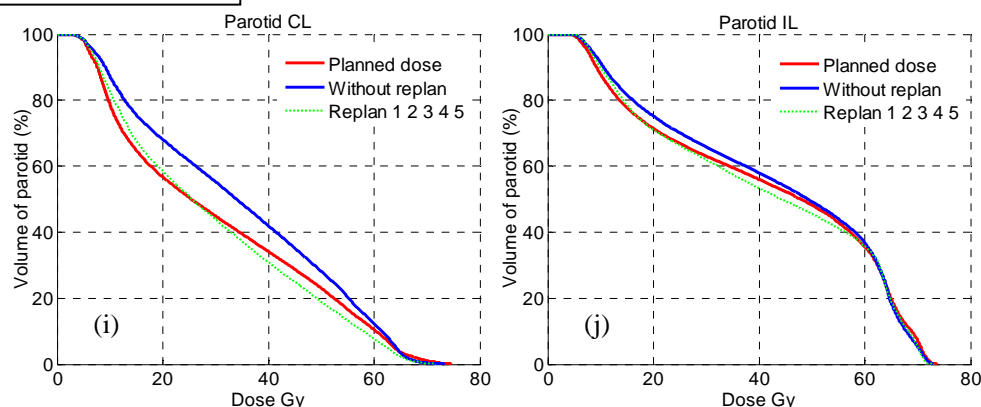


Figure 5.5. The DVH curves for the accumulated doses of scenarios with the different number and time of replannings. In the figure, from top to bottom, the DVH curves are represented for scenarios with the number of replannings from one to five times. The left column is the accumulated dose for contralateral (CL) parotid gland; the right column is the accumulated dose for ipsilateral parotid gland (IL).

The accumulated doses were analyzed by calculating the mean dose and NTCP values for each of the two parotid glands, as shown in Fig. 5.6.

It can be observed that, when only one replanning is performed, it should be done as soon as possible (even the first week of treatment). This can be explained by the delay between the planning and the starting of the treatment and by the fact that an early replanning has an impact on all the subsequent treatment fractions. It can also be observed that, when the number of replannings was increased, there was a larger probability for the mean dose of two parotid glands to be lower than the initial dose (represented by the red line). Thus, the more replannings used in scenarios, the more protection was given to the normal tissues.

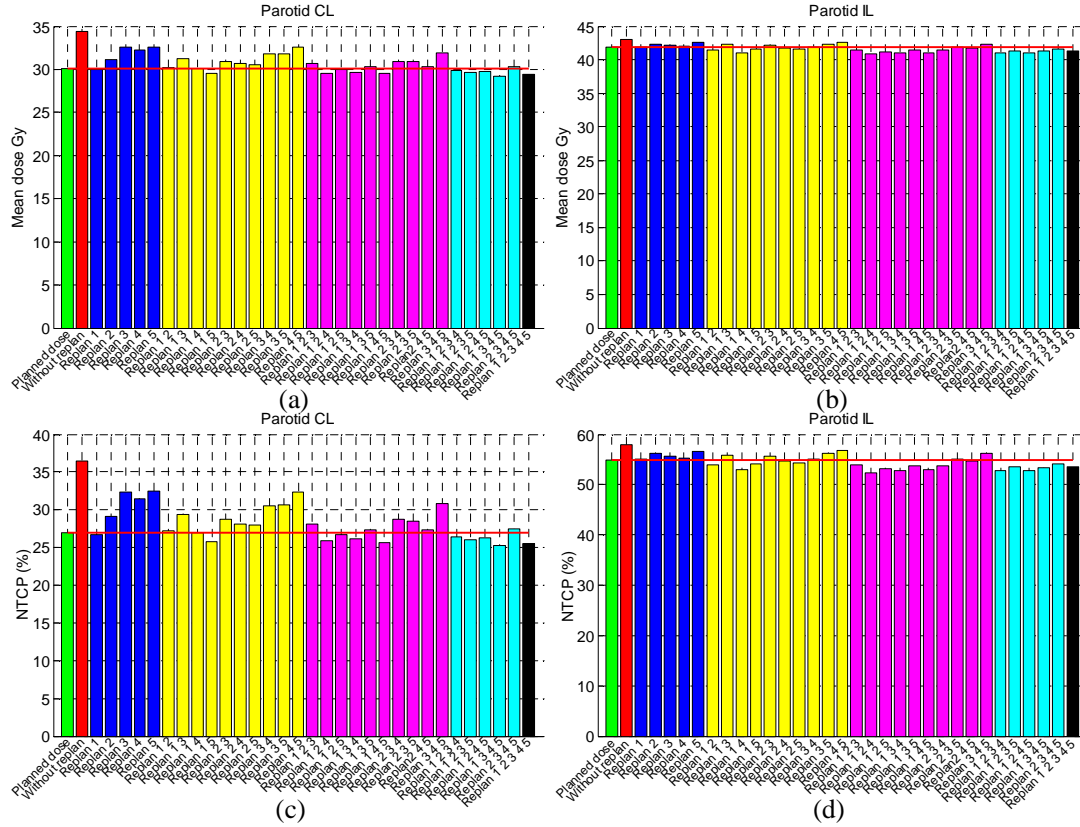


Figure 5.6. The mean dose and NTCP values for the accumulated dose of different scenarios: (a) the mean dose for contralateral parotid gland (CL); (b) the mean dose for ipsilateral parotid gland (IL); (c) NTCP for contralateral parotid gland; (d) NTCP for ipsilateral parotid gland. The green bar and the red line correspond to the planned dose.

V.5.4. Optimal number and time of replannings

- For both parotids, the best scenarios are shown in Table 5.2. The optimal combination was with three replannings at first, second and fourth weeks of treatment. Compared to the scenario without replanning, the optimal scenario decreased EUD by 3.53Gy (NTCP by 8.66%).

- For ipsilateral parotid glands (Table 5.3.), the optimal combination was with three replannings realized at a the first, second and fourth week of treatment. Compared to the scenario without replanning, the optimal scenario decreased the EUD by 2.26Gy (and the NTCP by 5.59%; $p=1.49E-06$).

- For contralateral parotid glands (Table 5.3.), the optimal combination was with four replannings realized at the first, third, fourth and fifth week of treatment. Compared to the scenario without replanning, the optimal scenario decreased the EUD by 5.14Gy (and the NTCP by 11.26%; $p= 4.89E-06$).

Optimal scenarios	EUD (Gy)	NTCP
Replan 1 2 4	35.21	38.45%
Replan 1 3 4	35.36	38.80%
Replan 1 4 5	35.29	38.62%
Replan 1 2 3 4	35.41	38.93%
Replan 1 2 4 5	35.41	38.92%
Replan 1 3 4 5	35.26	38.56%
Replan 1 2 3 4 5	35.35	38.78%

Table 5.2. Optimal scenarios of replanning for both parotid glands.

	Optimal scenarios	EUD (Gy)	NTCP	<i>p</i> value
IL	Replan 1 2 4	40.85	52.37%	1.49E-06
	Replan 1 3 4	41.02	52.79%	3.22E-05
	Replan 1 2 3 4	41.00	52.76%	2.73E-05
CL	Replan 1 5	29.51	25.74%	0.0001
	Replan 1 2 4	29.58	25.89%	0.0046
	Replan 1 4 5	29.49	25.71%	0.0006
	Replan 1 3 4 5	29.24	25.21%	4.89E-06
	Replan 1 2 3 4 5	29.41	25.55%	0.0001

Table 5.3. Optimal scenarios of replannings considering each side of the PGs.

V.5. Discussion

In this part of the work, we proposed a method to determine the optimal time and number of replannings for locally advanced head and neck cancer patients. Our proposed method calculated the accumulated dose for each scenario at first. Then, the optimal scenario was obtained by analyzing the dosimetric impacts of the cumulated dose for all the scenarios. The goal was to obtain the optimal number and time of replannings. Compared to the methods published in the literature^[102-104], our method took into account the influences of the current replanning on the overall dose distributions, rather than to a particular fractional irradiation. The optimal combination appears to be with three replannings realized at the first, second and fourth week of treatment. More replannings don't appear to increase the PG dose sparing. It could be interesting to check if the same results would be obtained with a reduced delay between the planning and the first week of treatment.

We reported the volume changes for ipsilateral parotid gland, contralateral parotid gland during the treatment. As reported in several papers^[19-22], the mean volume of both parotid glands decreased of around 30% (cf. Fig. 5.3).

We used a contour-guided Demons registration algorithm. Since PG shape and volume variations were limited, our study's Dice scores were relatively high (0.92). However, the Dice score does not provide any information regarding the registration's anatomical "point to point" correspondence accuracy.

V.6. Conclusion

In this part of work, we seek to identify the optimal adaptive RT scenario for the treatment of locally advanced head-and-neck cancer to spare the PG and limit the risk of xerostomia. The dosimetric criteria were based on the cumulated dose using non-rigid registration and dose warping. The optimal scenario was obtained by analyzing the dosimetric impacts of the different replanning scenarios on these cumulated doses in the PGs. Compared to the methods of the literature, our proposed method took into account the influences of the replannings to the overall dose distributions, rather than to a particular fractional irradiation. In our study, we calculated the optimal time and number of replannings.

Further works should be done to identify the anatomical descriptors enabling to trigger a replanning according to the anatomical evolution of each individual patient.

Chapter VI. Conclusions

This thesis mainly focused on the pencil beam dose calculation and plan optimization algorithms. Some methods were also proposed for the correction of random geometric uncertainties in IMRT, and to determine the optimal time and number of replannings in ART.

Firstly, a new pencil beam model was proposed to avoid the bias caused by oblique kernels and to take account for heterogeneities. To avoid any oblique kernel-related bias and reduce computation time, dose distributions were computed in a spherical coordinate system based on the pencil kernels of different distances from source to surface (DSS). We employed two different dose calculation methods: the superposition method and the fast Fourier transform convolution (FFTC) method. In order to improve the superposition method, we scaled the depth-directed component by moving the position of the entry point and altering the DSS value for a given beamlet. The lateral components were thus directly corrected by the density scaling method along the spherical shell without taking the densities from the previous layers into account. Significant computation time could be saved by performing the FFTC calculations on each spherical shell, disregarding density changes in the lateral direction. The proposed methods were tested on several phantoms, including lung- and bone-type heterogeneities and were compared with Monte Carlo (MC) simulation and AAA algorithm. Our proposed method achieved better dose calculation accuracy without increasing computational time compared to the AAA algorithm. The FFTC method was approximately 40 times faster than the superposition method, though this rapidity was at the cost of reduced dose calculation accuracy. This method could however be interesting when very fast dose calculations are needed, for example, in online ART workflows. Moreover, in a plan optimization strategy, it could be interesting to combine both methods: the FFTC dose calculation to grossly estimate the dose, combined with the superposition or MC method to estimate the dose when the algorithm starts to reach convergence. Further works will deal with evaluations realized on real patient images.

Secondly, a new method was proposed to incorporate the equivalent convex NTCP constraints into an optimization approach. We described the sub-score of an equivalent convex NTCP model [9], and provided its first derivatives. The efficacy of the equivalent convex NTCP constraints, along with their use in large-scale constrained optimization, was investigated. A gradient-based optimization algorithm, namely L-BFGS, was applied so as to solve this optimization problem. The effectiveness of the proposed method was assessed in 10 prostate cancer cases. The results proved that the proposed method was capable of generating clinically acceptable plans which meet the practical clinical guidelines. The proposed method is a FMO method, which improved quality of the two step approach in IMRT optimizations. For the convex sub-scores used in this work, these sub-scores could be directly applied to the direct aperture optimization, such as the Column-generation approach. Moreover, convexity of the sub-scores is also sufficient to guarantee that the Pareto efficient frontier is convex in case of the multi-criteria FMO

model. Future works may address the extension of the proposed physical-biological model to applications in direct aperture optimization.

Thirdly, a new deconvolution method was proposed to correct random geometrical uncertainties that are intrinsically observed in radiation therapy. The new approach using both a series expansion and a Butterworth filter was tested on two 2D fluence maps and a prostate case. It was shown that the Butterworth filter better suppressed high-frequency oscillations and reduced hot and cold spots on convolved-back fluence profiles. In the flat area of the fluence map, our method's accuracy was similar to that of the deconvolution kernel method. Near the edge of the fluence map, a clear advantage was observed. This method has been compared to the deconvolution kernel method for a regular 2D fluence map, a real intensity-modulated radiation therapy (IMRT) field, and a prostate case. The results showed that accuracy is improved, while fulfilling clinical planning requirements. The proposed approach determined the robust fluence from a static fluence and provided a straightforward understanding of the general feature of the robust fluence versus the static one. However, it could be interesting to simulate the dosimetric impact of this approach, for example by considering per-treatment images and in terms of cumulated dose.

Fourthly, a study was carried out to determine both the optimal time and number of replannings for locally advanced head and neck cancer patients. The optimal results were obtained by analyzing the accumulated doses, which took into account the influences of the current replanning to the overall dose distributions, rather than to the correction for a particular fractional irradiation. The optimal combination was obtained with three replannings realized at the first, second and fourth week of treatment. These results suggest to decrease the delay between the planning and the beginning of treatment. Moreover, relating the benefit of replanning with anatomical changes would be a great advance to effectively implement an ART workflow.

Thus, we presented in this work different methods for planning optimization in radiation therapy of prostate and head-and-neck cancer. These methods have been evaluated and enable to improve the planning in several aspect. Their combination in a global workflow should be performed to evaluate their actual clinical impact to improve the patient outcome.

References

- [1] Lemoigne Y, Caner A. Radiotherapy and Brachytherapy [M]. Berlin: Springer, 2009:11, 2009, pp11.
- [2] Jin C. The accurate radiotherapy biological effects of image-guided radiotherapy [D]; University of Science and Technology of China, 2012.
- [3] Dunne-Daly C F. Principles of radiotherapy and radiobiology [C]. Proc. of the Seminars in oncology nursing, 1999, 15(4):250-259.
- [4] Shupeng L, Liling W. Advances of precision radiotherapy [J]. Chinese Medical Equipment Journal, 2007, 28(11): 57-58.
- [5] Determination of absorbed dose in a patient irradiated by beams of X or Gamma rays in radiotherapy. International commission on radiation units and measurements (ICRU), ICRU Report No. 24, 1976, Bethesda, Maryland.
- [6] Boyer A L, Schultheiss T. Effects of dosimetric and clinical uncertainty on complication-free local tumor control [J]. Radiotherapy and Oncology, 1988, 11(1): 65-71.
- [7] Siebers J V, Tong S, Lauterbach M, et al. Acceleration of dose calculations for intensity-modulated radiotherapy [J]. Medical physics, 2001, 28(6): 903-910.
- [8] Brahme A. Radiation Therapy Physics [M]. Berlin: Springer. 1995, pp209-246.
- [9] Hoffmann A L, den Hertog D, Siem A Y, et al. Convex reformulation of biologically-based multi-criteria intensity-modulated radiation therapy optimization including fractionation effects [J]. Physics in medicine and biology, 2008, 53(22): 6345-6362.
- [10] Stroom J C, Heijmen B J. Geometrical uncertainties, radiotherapy planning margins, and the ICRU-62 report [J]. Radiotherapy and oncology, 2002, 64(1): 75-83.
- [11] Van Herk M. Errors and margins in radiotherapy [C]. Proc. of the Seminars in radiation oncology, 2004, 14(1):52-64.
- [12] Yan D, Ziaja E, Jaffray D, et al. The use of adaptive radiation therapy to reduce setup error: a prospective clinical study [J]. International Journal of Radiation Oncology Biology Physics, 1998, 41(3): 715-720.
- [13] Hoogeman M S, van Herk M, de Bois J, et al. Strategies to reduce the systematic error due to tumor and rectum motion in radiotherapy of prostate cancer [J]. Radiotherapy and oncology, 2005, 74(2): 177-185.
- [14] de la Zerda A, Armbruster B, Xing L. Formulating adaptive radiation therapy (ART) treatment planning into a closed-loop control framework [J]. Physics in medicine and biology, 2007, 52(14): 4137-4153.
- [15] Peng C, Chen G, Ahunbay E, et al. Validation of an online replanning technique for prostate adaptive radiotherapy [J]. Physics in medicine and biology, 2011, 56(12): 3659-3668.
- [16] Suh Y, Sawant A, Venkat R, et al. Four-dimensional IMRT treatment planning using a DMLC motion-tracking algorithm [J]. Physics in medicine and biology, 2009, 54(12): 3821-3835.
- [17] Fan Y, Nath R. Intensity modulation under geometrical uncertainty: a deconvolution approach to robust fluence [J]. Physics in medicine and biology, 2010, 55(14): 4029-4045.
- [18] Webb S. Quantification of the fluence error in the motion-compensated dynamic MLC (DMLC) technique for delivering intensity-modulated radiotherapy (IMRT) [J]. Physics in medicine and biology, 2006, 51(7): L17-L21.
- [19] Castadot P, Lee J A, Geets X, et al. Adaptive radiotherapy of head-and-neck cancer [C]. Proc. of the Seminars in radiation oncology, 2010, 20(2):84-93.
- [20] Schwartz D L, Dong L. Adaptive radiation therapy for head-and-neck cancer—can an old goal evolve into a new standard? [J]. Journal of oncology, 2010, 2011:1-13.
- [21] Wu Q, Chi Y, Chen P Y, et al. Adaptive replanning strategies accounting for shrinkage in

-
- head-and-neck IMRT [J]. *International Journal of Radiation Oncology Biology Physics*, 2009, 75(3): 924-932.
- [22] Barker Jr J L, Garden A S, Ang K K, et al. Quantification of volumetric and geometric changes occurring during fractionated radiotherapy for head-and-neck cancer using an integrated CT/linear accelerator system [J]. *International Journal of Radiation Oncology Biology Physics*, 2004, 59(4): 960-970.
- [23] Han C, Chen Y-J, Liu A, et al. Actual dose variation of parotid glands and spinal cord for nasopharyngeal cancer patients during radiotherapy [J]. *International Journal of Radiation Oncology Biology Physics*, 2008, 70(4): 1256-1262.
- [24] Lee C, Langen K M, Lu W, et al. Assessment of parotid gland dose changes during head-and-neck cancer radiotherapy using daily megavoltage computed tomography and deformable image registration [J]. *International Journal of Radiation Oncology Biology Physics*, 2008, 71(5): 1563-1571.
- [25] Ulmer W, Pyry J, Kaissl W. A 3D photon superposition/convolution algorithm and its foundation on results of Monte Carlo calculations [J]. *Physics in medicine and biology*, 2005, 50(8): 1767-1790.
- [26] Tillikainen L, Helminen H, Torsti T, et al. A 3D pencil-beam-based superposition algorithm for photon dose calculation in heterogeneous media [J]. *Physics in medicine and biology*, 2008, 53(14): 3821-3829.
- [27] Lu W. A non-voxel-based broad-beam (NVBB) framework for IMRT treatment planning [J]. *Physics in medicine and biology*, 2010, 55(23): 7175-7210.
- [28] Lu W, Chen M. Fluence-convolution broad-beam (FCBB) dose calculation [J]. *Physics in medicine and biology*, 2010, 55(23): 7211-7229.
- [29] Ahnesjö A, Saxner M, Trepp A. A pencil beam model for photon dose calculation [J]. *Medical physics*, 1992, 19(2): 263-273.
- [30] Storchi P, Van Battum L, Woudstra E. Calculation of a pencil beam kernel from measured photon beam data [J]. *Physics in medicine and biology*, 1999, 44(12): 2917-2928.
- [31] Jeleń U, Söhn M, Alber M. A finite size pencil beam for IMRT dose optimization [J]. *Physics in medicine and biology*, 2005, 50(8): 1747-1766.
- [32] Jeleń U, Alber M. A finite size pencil beam algorithm for IMRT dose optimization: density corrections [J]. *Physics in medicine and biology*, 2007, 52(3): 617-633.
- [33] Siggel M, Ziegenhein P, Nill S, et al. Boosting runtime-performance of photon pencil beam algorithms for radiotherapy treatment planning [J]. *Physica Medica*, 2012, 28(4): 273-280.
- [34] Cilla S, Digesù C, Macchia G, et al. Clinical implications of different calculation algorithms in breast radiotherapy: A comparison between pencil beam and collapsed cone convolution [J]. *Physica Medica*, 2014, 30(4): 473-481.
- [35] Zhu Y, Boyer A. X-ray dose computations in heterogeneous media using 3-dimensional FFT convolution [J]. *Physics in Medicine and Biology*, 1990, 35(3): 351-368.
- [36] Erdi A K, Yorke E D, Loew M H, et al. Use of the fast Hartley transform for three-dimensional dose calculation in radionuclide therapy [J]. *Medical physics*, 1998, 25(11): 2226-2233.
- [37] Ahnesjö A, Aspradakis M M. Dose calculations for external photon beams in radiotherapy [J]. *Physics in medicine and biology*, 1999, 44(11): R99-R155.
- [38] Sharpe M, Battista J. Dose calculations using convolution and superposition principles: The orientation of dose spread kernels in divergent x - ray beams [J]. *Medical physics*, 1993, 20(6): 1685-1694.
- [39] Knöös T, Wieslander E, Cozzi L, et al. Comparison of dose calculation algorithms for treatment planning in external photon beam therapy for clinical situations [J]. *Physics in medicine and biology*, 2006, 51(22): 5785-5807.
- [40] Kawrakow I. Accurate condensed history Monte Carlo simulation of electron transport. I. EGSnrc, the new EGS4 version [J]. *Medical physics*, 2000, 27(3): 485-498.

-
- [41] Seco J, Evans P. Assessing the effect of electron density in photon dose calculations [J]. *Medical physics*, 2006, 33(2): 540-552.
- [42] Ahnesjö A. Collapsed cone convolution of radiant energy for photon dose calculation in heterogeneous media [J]. *Medical physics*, 1989, 16(4): 577-592.
- [43] Mohan R, Chui C, Lidofsky L. Energy and angular distributions of photons from medical linear accelerators [J]. *Medical physics*, 1985, 12(5): 592-597.
- [44] Rogers D, Walters B, Kawrakow I. BEAMnrc users manual [J]. NRC Report PIRS, 2001.
- [45] Walters B, Kawrakow I, Rogers D. DOSXYZnrc users manual [J]. NRC Report PIRS, 2005.
- [46] Miften M, Wiesmeyer M, Monthofer S, et al. Implementation of FFT convolution and multigrid superposition models in the FOCUS RTP system [J]. *Physics in medicine and biology*, 2000, 45(4): 817-833.
- [47] Fogliata A, Vanetti E, Albers D, et al. On the dosimetric behaviour of photon dose calculation algorithms in the presence of simple geometric heterogeneities: comparison with Monte Carlo calculations [J]. *Physics in medicine and biology*, 2007, 52(5): 1363-1385.
- [48] Siebers J V, Lauterbach M, Tong S, et al. Reducing dose calculation time for accurate iterative IMRT planning [J]. *Medical physics*, 2002, 29(2): 231-237.
- [49] Siebers J V, Kawrakow I, Ramakrishnan V. Performance of a hybrid MC dose algorithm for IMRT optimization dose evaluation [J]. *Medical physics*, 2007, 34(7): 2853-2863.
- [50] Wu Q, Mohan R, Niemierko A, et al. Optimization of intensity-modulated radiotherapy plans based on the equivalent uniform dose [J]. *International Journal of Radiation Oncology Biology Physics*, 2002, 52(1): 224-235.
- [51] Bortfeld T, Neve W, Schmidt-Ullrich R, et al. Image-guided IMRT [M]. Springer, 2006.
- [52] Choi B, Deasy J O. The generalized equivalent uniform dose function as a basis for intensity-modulated treatment planning [J]. *Physics in medicine and biology*, 2002, 47(20): 3579-3589.
- [53] Das S. A role for biological optimization within the current treatment planning paradigm [J]. *Medical physics*, 2009, 36(10): 4672-4682.
- [54] Dirscherl T, Alvarez-Moret J, Bogner L. Advantage of biological over physical optimization in prostate cancer? [J]. *Zeitschrift für Medizinische Physik*, 2011, 21(3): 228-235.
- [55] Diot Q, Kavanagh B, Timmerman R, et al. Biological-based optimization and volumetric modulated arc therapy delivery for stereotactic body radiation therapy [J]. *Medical physics*, 2011, 39(1): 237-245.
- [56] Qi X S, Semenenko V A, Li X A. Improved critical structure sparing with biologically based IMRT optimization [J]. *Medical physics*, 2009, 36(5): 1790-1799.
- [57] Li X A, Alber M, Deasy J O, et al. The use and QA of biologically related models for treatment planning: Short report of the TG-166 of the therapy physics committee of the AAPM [J]. *Medical physics*, 2012, 39(3): 1386-1409.
- [58] Mohan R, Mageras G, Baldwin B, et al. Clinically relevant optimization of 3-D conformal treatments [J]. *Medical physics*, 1992, 19(4): 933-944.
- [59] Stavrev P, Hristov D, Warkentin B, et al. Inverse treatment planning by physically constrained minimization of a biological objective function [J]. *Medical physics*, 2003, 30(11): 2948-2958.
- [60] Kessler M L, Mcshan D L, Epelman M A, et al. Costlets: A generalized approach to cost functions for automated optimization of IMRT treatment plans [J]. *Optimization and Engineering*, 2005, 6(4): 421-448.
- [61] Wang X-H, Mohan R, Jackson A, et al. Optimization of intensity-modulated 3D conformal treatment plans based on biological indices [J]. *Radiotherapy and Oncology*, 1995, 37(2): 140-152.
- [62] Jeraj R, Wu C, Mackie T R. Optimizer convergence and local minima errors and their clinical importance [J]. *Physics in medicine and biology*, 2003, 48(17): 2809-2827.
- [63] Hindi H. A tutorial on convex optimization II: Duality and interior point methods [C]. *Proc.*

-
- of the American Control Conf, Boston: IEEE, 2004, 4: pp 3252–65.
- [64] Romeijn H E, Dempsey J F, Li J G. A unifying framework for multi-criteria fluence map optimization models [J]. *Physics in medicine and biology*, 2004, 49(10): 1991-2013.
- [65] Liu D C, Nocedal J. On the limited memory BFGS method for large scale optimization [J]. *Mathematical programming*, 1989, 45(1-3): 503-528.
- [66] Lyman J T. Complication probability as assessed from dose-volume histograms [J]. *Radiation Research*, 1985, 104(2s): S13-S19.
- [67] Kutcher G J, Burman C. Calculation of complication probability factors for non-uniform normal tissue irradiation: The effective volume method gerald [J]. *International Journal of Radiation Oncology Biology Physics*, 1989, 16(6): 1623-1630.
- [68] Niemierko A. A unified model of tissue response to radiation (abstract) [J]. *Medical physics*, 1999, 26(6): 1100.
- [69] Wu Q, Mohan R. Algorithms and functionality of an intensity modulated radiotherapy optimization system [J]. *Medical physics*, 2000, 27(4): 701-711.
- [70] Deasy J O, Blanco A I, Clark V H. CERR: a computational environment for radiotherapy research [J]. *Medical physics*, 2003, 30(5): 979-985.
- [71] Dale E, Hellebust T P, Skjærnsberg A, et al. Modeling normal tissue complication probability from repetitive computed tomography scans during fractionated high-dose-rate brachytherapy and external beam radiotherapy of the uterine cervix [J]. *International Journal of Radiation Oncology Biology Physics*, 2000, 47(4): 963-971.
- [72] Peeters S T, Hoogeman M S, Heemsbergen W D, et al. Rectal bleeding, fecal incontinence, and high stool frequency after conformal radiotherapy for prostate cancer: normal tissue complication probability modeling [J]. *International Journal of Radiation Oncology Biology Physics*, 2006, 66(1): 11-19.
- [73] Mageras G, Mohan R. Application of fast simulated annealing to optimization of conformal radiation treatments [J]. *Medical physics*, 1993, 20(3): 639-647.
- [74] Marks L B, Yorke E D, Jackson A, et al. Use of normal tissue complication probability models in the clinic [J]. *International Journal of Radiation Oncology Biology Physics*, 2010, 76(3): S10-S19.
- [75] Emami B, Lyman J, Brown A, et al. Tolerance of normal tissue to therapeutic irradiation [J]. *International Journal of Radiation Oncology Biology Physics*, 1991, 21(1): 109-122.
- [76] Burman C, Kutcher G, Emami B, et al. Fitting of normal tissue tolerance data to an analytic function [J]. *International Journal of Radiation Oncology Biology Physics*, 1991, 21(1): 123-135.
- [77] Fiorino C, Valdagni R, Rancati T, et al. Dose–volume effects for normal tissues in external radiotherapy: pelvis [J]. *Radiotherapy and Oncology*, 2009, 93(2): 153-167.
- [78] Fellin G, Fiorino C, Rancati T, et al. Late rectal bleeding after conformal radiotherapy for prostate cancer: NTCP Modeling [J]. *Radiother Oncol*, 2008, 72(1): S332-S333.
- [79] Söhn M, Yan D, Liang J, et al. Incidence of late rectal bleeding in high-dose conformal radiotherapy of prostate cancer using equivalent uniform dose–based and dose–volume–based normal tissue complication probability models [J]. *International Journal of Radiation Oncology Biology Physics*, 2007, 67(4): 1066-1073.
- [80] Tucker S, Dong L, Bosch W, et al. Fit of a generalized Lyman Normal tissue complication probability (NTCP) model to grade \geq 2 late rectal toxicity data from patients treated on protocol RTOG 94-06 [J]. *Int J Radiat Oncol Biol Phys* 2007, 69(1): S8- S9.
- [81] Rancati T, Fiorino C, Gagliardi G, et al. Fitting late rectal bleeding data using different NTCP models: results from an Italian multi-centric study (AIROPROS0101) [J]. *Radiotherapy and oncology*, 2004, 73(1): 21-32.
- [82] Baum C, Alber M, Birkner M, et al. Robust treatment planning for intensity modulated radiotherapy of prostate cancer based on coverage probabilities [J]. *Radiotherapy and oncology*, 2006, 78(1): 27-35.
- [83] Lujan A E, Ten Haken R K, Larsen E W, et al. Quantization of setup uncertainties in 3-D

-
- dose calculations [J]. *Medical physics*, 1999, 26(11): 2397-2402.
- [84] McCarter S, Beckham W. Evaluation of the validity of a convolution method for incorporating tumour movement and set-up variations into the radiotherapy treatment planning system [J]. *Physics in medicine and biology*, 2000, 45(4): 923-931.
- [85] Craig T, Battista J, Van Dyk J. Limitations of a convolution method for modeling geometric uncertainties in radiation therapy. I. The effect of shift invariance [J]. *Medical physics*, 2003, 30(8): 2001-2011.
- [86] Lu W, Olivera G H, Mackie T R. Motion-encoded dose calculation through fluence/sinogram modification [J]. *Medical physics*, 2004, 32(1): 118-127.
- [87] Lind B K, Källman P, Sundelin B, et al. Optimal radiation beam profiles considering uncertainties in beam patient alignment [J]. *Acta Oncologica*, 1993, 32(3): 331-342.
- [88] Lof J, Lind B K, Brahme A. Optimal radiation beam profiles considering the stochastic process of patient positioning in fractionated radiation therapy [J]. *Inverse Problems*, 1995, 11(6): 1189-1209.
- [89] Gordon J, Siebers J. Evaluation of dosimetric margins in prostate IMRT treatment plans [J]. *Medical physics*, 2008, 35(2): 569-575.
- [90] Bortfeld T, Chan T C, Trofimov A, et al. Robust management of motion uncertainty in intensity-modulated radiation therapy [J]. *Operations Research*, 2008, 56(6): 1461-1473.
- [91] Ulmer W, Kaissl W. The inverse problem of a Gaussian convolution and its application to the finite size of the measurement chambers/detectors in photon and proton dosimetry [J]. *Physics in medicine and biology*, 2003, 48(6): 707-727.
- [92] Ulmer W. Inverse problem of linear combinations of Gaussian convolution kernels (deconvolution) and some applications to proton/photon dosimetry and image processing [J]. *Inverse Problems*, 2010, 26(8): 1-26.
- [93] Fang T-M, Shei S-S, Nagem R, et al. Convolution and deconvolution with Gaussian kernel [J]. *Il Nuovo Cimento B Series 11*, 1994, 109(1): 83-92.
- [94] Garcia-Vicente F, Delgado J, Rodriguez C. Exact analytical solution of the convolution integral equation for a general profile fitting function and Gaussian detector kernel [J]. *Physics in medicine and biology*, 2000, 45(3): 645-650.
- [95] Moore J A, Gordon J J, Anscher M S, et al. Comparisons of treatment optimization directly incorporating random patient setup uncertainty with a margin-based approach [J]. *Medical physics*, 2009, 36(9): 3880-3890.
- [96] van Herk M, Remeijer P, Lebesque J V. Inclusion of geometric uncertainties in treatment plan evaluation [J]. *International Journal of Radiation Oncology Biology Physics*, 2002, 52(5): 1407-1422.
- [97] Low D A, Harms W B, Mutic S, et al. A technique for the quantitative evaluation of dose distributions [J]. *Medical physics*, 1998, 25(5): 656-661.
- [98] Engel K. A new algorithm for optimal multileaf collimator field segmentation [J]. *Discrete Applied Mathematics*, 2005, 152(1): 35-51.
- [99] Vásquez Osorio E M, Hoogeman M S, Al-Mamgani A, et al. Local anatomic changes in parotid and submandibular glands during radiotherapy for oropharynx cancer and correlation with dose, studied in detail with nonrigid registration [J]. *International Journal of Radiation Oncology Biology Physics*, 2008, 70(3): 875-882.
- [100] Geets X, Tomsej M, Lee J A, et al. Adaptive biological image-guided IMRT with anatomic and functional imaging in pharyngo-laryngeal tumors: impact on target volume delineation and dose distribution using helical tomotherapy [J]. *Radiotherapy and Oncology*, 2007, 85(1): 105-115.
- [101] Robar J L, Day A, Clancey J, et al. Spatial and dosimetric variability of organs at risk in head-and-neck intensity-modulated radiotherapy [J]. *International Journal of Radiation Oncology Biology Physics*, 2007, 68(4): 1121-1130.
- [102] Ahamad A, Dong L, Zhang L, et al. Is there a trigger point for adaptive replanning during

-
- head & neck IMRT? [J]. *International Journal of Radiation Oncology Biology Physics*, 2006, 66(3): S100-S101.
- [103] Meldolesi E, Wu Q, Chen P, et al. Evaluation of anatomic and dosimetric changes during treatment course of head-and-neck (HN) IMRT: Is replanning necessary? [J]. *International Journal of Radiation Oncology Biology Physics*, 2006, 66(3): S101-S102.
- [104] Siddiqui F, Boyle D, Weldon M, et al. Adaptive radiation therapy in head-and-neck Cancers: When should we re-plan? [J]. *International Journal of Radiation Oncology Biology Physics*, 2011, 81(2): S494-S495.
- [105] Wu Q, Nandalur S R, Nandalur K R, et al. What time is optimal for replanning head-and-neck IMRT (HN-IMRT)? [J]. *International Journal of Radiation Oncology Biology Physics*, 2008, 72(1): S163-S163.
- [106] Vercauteren T, Pennec X, Perchant A, et al. Symmetric log-domain diffeomorphic registration: A demons-based approach [M]. *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2008*. Springer. 2008: 754-761.
- [107] Thirion J-P. Image matching as a diffusion process: an analogy with Maxwell's demons [J]. *Medical image analysis*, 1998, 2(3): 243-260.
- [108] Dijkema T, Raaijmakers C P, Ten Haken R K, et al. Parotid gland function after radiotherapy: the combined Michigan and Utrecht experience [J]. *International Journal of Radiation Oncology Biology Physics*, 2010, 78(2): 449-453.

Publications

Journals

[1] Pengcheng Zhang, Jian Zhu, Yuanjin Li, Huazhong Shu. Improved algorithm of multileaf collimator field segmentation. Journal of southeast university (Natural Science Edition), 2012, 42(5):875-879. (Chinese)

[2] Pengcheng Zhang, Renaud De Crevoisier, Antoine Simon, Pascal Haigron, Jean-Louis Coatrieux, Baosheng Li and Huazhong Shu. A new deconvolution approach to robust fluence for intensity modulation under geometrical uncertainty. Phys Med Biol, 2013, 58(17):6095-6110.

[3] Pengcheng Zhang, Antoine Simon, Renaud De Crevoisier, Pascal Haigron, Mohamed Hatem Nassef, Baosheng Li and Huazhong Shu. A new pencil beam model for photon dose calculations in heterogeneous media, Physica Medica, accepted.

[4] Pengcheng Zhang, Antoine Simon, Huazhong Shu, Pascal Haigron, Caroline Lafond, Jean-Louis Coatrieux, Jian Zhu, Baosheng Li and Renaud De Crevoisier. IMRT optimization including the equivalent convex NTCP constraint, submitted.

Conference

[1] Pengcheng Zhang, Huazhong Shu, Antoine Simon, Pascal Haigron and Renaud De Crevoisier. A new pencil beam model for photon dose calculations. The 2013 6th International Conference on BioMedical Engineering and Informatics (BMEI 2013)

Manuscrit en Chinois

东南大学学位论文独创性声明

本人声明所呈交的学位论文是我个人在导师指导下进行的研究工作及取得的研究成果。尽我所知，除了文中特别加以标注和致谢的地方外，论文中不包含其他人已经发表或撰写过的研究成果，也不包含为获得东南大学或其它教育机构的学位或证书而使用过的材料。与我一同工作的同志对本研究所做的任何贡献均已在论文中作了明确的说明并表示了谢意。

研究生签名：_____ 日期：_____

东南大学学位论文使用授权声明

东南大学、中国科学技术信息研究所、国家图书馆有权保留本人所送交学位论文的复印件和电子文档，可以采用影印、缩印或其他复制手段保存论文。本人电子文档的内容和纸质论文的内容相一致。除在保密期内的保密论文外，允许论文被查阅和借阅，可以公布（包括刊登）论文的全部或部分内容。论文的公布（包括刊登）授权东南大学研究生院办理。

研究生签名：_____ 导师签名：_____ 日期：_____

摘要

放射治疗是治疗肿瘤的主要方法之一。在放射治疗系统中进行精确计划是实现精确放射治疗的重要环节，而精确的放射剂量计算方法及快速、有效的调强优化算法是制定高质量放疗方案的前提和保障。

本文主要致力于笔形束剂量计算方法和方案优化算法的研究，在总结现有放疗剂量计算方法及方案优化算法的基础上，对已有的经典笔形束算法进行改进，增加剂量计算的精度；在调强放射治疗中，将等效凸生物准则应用到方案优化的目标函数中，保证方案优化结果为全局生物最优。在上述工作的基础上，我们进一步研究了调强放射治疗中随机几何不确定性的校正问题以及自适应放射治疗中最佳重新规划时间和次数的问题。

在调强放射治疗过程中，笔形束算法经常被用来计算剂量分布。在第一部分中，提出了一种新的笔形束模型来计算非均匀介质中的光子剂量分布。为了校正由倾斜的笔形束核所引入的计算误差并缩短校正过程所花费的计算时间，使用不同射线源到皮肤表层距离(distance from source to surface, DSS)的笔形束核在球壳坐标系统中计算剂量分布，并分别采用叠加算法和快速傅立叶变换算法计算剂量分布。为了增加叠加剂量计算方法的计算精度，通过移动射束在体模表层入射点的位置和修改对应射束笔形束核的 DSS，对深度分量进行校正，从而直接使用密度拉伸算法在当前球壳上对剖面分量进行校正，而无需考虑不同深度的球壳上的密度信息。在不考虑侧向非均匀介质分布的情况下，可以在每层球壳上直接使用快速傅立叶变换算法对剂量计算进行加速，减少了剂量计算所需时间。实验结果表明，叠加剂量计算方法的精度高于经典笔形束算法；FFTC 方法的精度较低，但是相比叠加剂量计算方法，计算速度提升了近 40 倍。

正常组织并发症概率(normal tissue complication probability, NTCP)准则与临床治疗密切相关，它可以直接估算在任何剂量分布下正常组织出现并发症的概率。但是，在引入 NTCP 这一生物因子参与调强放疗注量图优化的过程中，NTCP 准则是关于注量分布的非线性、非凸函数。逆向计划中，直接在目标函数中使用 NTCP 准则增加了求解优化问题的难度。在第二部分中，我们将等效凸 NTCP 准则应用到逆向方案优化中，并与其它方案优化算法进行了比较。在逆向计划中，我们使用基于物理-生物准则的混合目标函数，其中等效凸 NTCP 准则被用来减少正常组织所接收到的剂量照射，物理准则被用来控制靶区剂量的分布。总目标函数是基于 NTCP 准则的子目标函数和基于物理准则的子目标函数的加权和。采用基于梯度的优化算法，即 limited-memory Broyden-Fletcher-Goldfarb-Shanno 算法，来求解大规模受限最小化优化问题。新算法还与基于剂量-体积准则的优化算法以及其它两种基于 NTCP 准则的优化算法在 10 例前列腺肿瘤患者上进行了对比实验。实验结果表明，利用新算法可以得到临床可接受的放射治疗计划，且相比其它两种优化算法，新算法所得放射治疗计划质量更高，所需计算时间更短。

由于随机误差具有有限的反应时间及不规则的运动形式，放射治疗中固有的随机几何不确定性不能被完全消除。在第三部分中，提出一种新的反卷积算法来校正随机几何不确定性。新算法由序列展开方法和巴特沃斯滤波器构成。为了抑制反卷积结果中的高频分量，只取序列展开式的前 4 项。根据不同的随机几何不确定性的概率密度函数调整巴特沃斯滤波器的参数，进一步滤除反卷积结果中的高频分量来缩小静态注量图和卷积后注量图的差异，特别是在射野的边缘位置处。为了减少靶区外正常组织接收到的剂量照射，鲁棒注量图在射野外的注量分布被置为零。实验结果表明，新算法提升了放射治疗的质量且满足临床放射治疗的要求。

放射治疗过程通常持续一个月之久，期间需要根据患者及肿瘤变化情况重新制定放射治疗方案。完全重新规划的工作量非常大，包括重新进行 CT 扫描、器官勾画和计划设计等。由于医疗资源有限，不可能为每个病人在每次分次照射之前进行重新规划。但是，可以使用有限次数的重新规划，最大程度地改善病人的放射治疗质量。在第四部分中，主要研究了离

线自适应放射治疗算法，提出了一种确定离线自适应放射治疗中最佳重新规划时间及次数的算法。相对于直接采用体模内解剖结构变化或剂量变化确定最佳重新规划时间的方法，新算法根据累积剂量分布来确定最佳重新规划的时间及次数，综合地考虑了重新规划对整体剂量分布的影响，而不仅仅是对某次分次照射进行校正。根据是否考虑腮腺组织的过剂量照射，分别给出了两种情况下头颈部肿瘤最佳重新规划的时间及次数。

关键词：剂量计算，笔形束，调强放射治疗，NTCP，自适应放射治疗，几何不确定性，反卷积

ABSTRACT

Radiotherapy is one of the main methods for tumor treatment. The plan generation in the radiotherapy treatment planning system plays an important role in the implementation of a precise radiotherapy treatment, which is highly depended on the dose calculation algorithms as well as the fast and effective plan optimization algorithms.

This thesis mainly focused on the pencil beam dose calculation and plan optimization algorithms. Based on the existing studies of the pencil beam dose calculation and plan optimization algorithms, the standard pencil beam algorithm was improved to increase the accuracy of dose calculations; the equivalent convex biological criteria were incorporated into the objective function in intensity modulated radiation therapy (IMRT) optimization, in order to ensure that the plan optimization achieves the global bio-optimal results. Based on the previous work, the relevant works were further investigated for the correction of the random geometric uncertainty in IMRT, and the optimal time and number of replannings in the adaptive radiotherapy.

The pencil beam method is frequently used for dose calculations in IMRT. In the first part of the dissertation, a new pencil model was proposed for photon dose calculations in heterogeneous media. To avoid the bias caused by oblique kernels and reduce the computation time, dose distributions was calculated in a spherical coordinate system with the pencil kernels of different distance from source to surface (DSS). Two methods have been used to calculate dose distributions: superposition method and fast Fourier transform convolution (FFTC) method. To improve the accuracy of the superposition method, the scaling of depth-directed components were modeled by moving the position of the entry point and altering the value of DSS for a given beamlet. Thus, the lateral components were directly corrected by the density scaling method along the spherical shell without considering the densities in the previous layers. Significant computation time savings could be gained by performing the FFTC calculations on each spherical shell without considering the density changes in the lateral direction. Results show that the superposition method is more accurate than the classical pencil beam method. The FFTC method is less accurate but about 40 times faster than the superposition method.

Normal tissue complication probability (NTCP) models are clinically relevant, since they estimate, by definition, the toxicity of normal tissue for any given dose distributions. However, NTCP criteria are a non-linear, non-convex model concerning fluence elements for fluence map optimization. Direct incorporation of such non-linear and non-convex criteria into objective function, renders a global optimum extremely difficult to find. In the second part, an equivalent convex NTCP constraint was incorporated into inverse planning optimization and its performance was compared to that of other optimization reference methods. A hybrid physical-biological model was proposed for inverse planning, including the application of equivalent convex NTCP criteria in order to minimize the dose delivered to normal tissue, and physical constraints were imposed to achieve the target dose. The objective function was defined as the sum of weighted NTCP-based and dose-based sub-scores. A gradient-based method, namely the limited-memory Broyden-Fletcher-Goldfarb-Shanno algorithm, was applied to resolve the issue of large-scale constrained minimization. The proposed method was compared to dose-volume-based optimization and the two other NTCP-based optimizations in 10 prostate cancer cases. The comparative studies showed that the proposed scheme was capable of generating clinically acceptable plans, and achieved higher quality treatment plans and required less computation time than the two other optimization approaches.

Random geometric uncertainties which are intrinsically observed in radiation therapy cannot

be fully eliminated due to the finite response time and irregular motion patterns. In the third part, random geometrical uncertainties were addressed by means of a new deconvolution method combining a series expansion and a Butterworth filter. In order to suppress the high-frequency components, only the first four terms of the series expansion were considered. By adjusting the parameters of the Butterworth filter according to the different probability density functions, high-frequency signal components were further suppressed to minimize differences between the static and convolved-back profiles, especially on the field edges. To reduce the dose delivered outside the field, a further approximation was made by setting the values of deconvolution profiles outside the field to zero. The results showed that accuracy was improved while fulfilling clinical planning requirements.

It always takes more than one month for the radiotherapy implementation, during which the plan needs to be adjusted according to the anatomical changes in the patients. Replanning is a heavy task, which includes getting CT scans, delineating organs, making a plan and so on. Due to the limited resources in hospital, it is impossible to replan a patient right before the fractional irradiation for each patient. However, within the clinically acceptable range, the number of replannings can be reduced properly. In the fourth part, a new algorithm was proposed to determine both the optimal time and number of replannings in the off-line adaptive radiotherapy. Compared to the methods which determine the optimal time of replannings according to the largest geometric or dosimetric changes during the radiotherapy, the new algorithm is based on the cumulative doses, which takes into account the influences of the current replanning to the overall dose distributions, rather than to the correction for a particular fractional irradiation. Two kinds of optimal results were provided depending on whether over-irradiated parotid glands were considered.

Key words: dose calculations, pencil beam, intensity modulated radiation therapy, NTCP, adaptive radiotherapy, geometric uncertainties, deconvolution.

目录

摘要.....	I
ABSTRACT.....	III
目录.....	V
本文常用略语说明.....	VIII
第一章 绪论.....	1
1.1 课题研究的背景和意义.....	1
1.2 国内外研究现状.....	2
1.2.1 光子剂量计算方法.....	2
1.2.2 IMRT 方案优化算法.....	6
1.2.3 IMRT 中几何不确定性的校正.....	7
1.2.4 自适应放射治疗.....	8
1.3 本文工作内容.....	9
1.4 参考文献.....	10
第二章 一种新的笔形束光子剂量计算方法.....	17
2.1 引言.....	17
2.2 笔形束剂量计算方法.....	18
2.3 新的笔形束剂量计算方法.....	19
2.3.1 获取笔形束核参数.....	19
2.3.2 叠加剂量计算方法.....	22
2.3.3 使用 FFTC 技术加速剂量计算.....	24
2.4 实验材料.....	24
2.5 实验结果.....	25
2.5.1 水体模.....	25
2.5.2 肺挡块体模.....	27
2.5.3 骨挡块体模.....	28
2.5.4 肺阻块体模和骨阻块体模.....	28
2.5.5 与 AAA 算法比较.....	29
2.5.6 计算时间.....	31
2.6 讨论.....	32
2.7 本章小结.....	32
2.8 参考文献.....	33
第三章 基于生物准则的 IMRT 方案优化算法的研究.....	35
3.1 引言.....	35
3.2 优化准则.....	36
3.2.1 物理准则.....	36
3.2.2 生物准则.....	36
3.3 使用生物准则的优点.....	41
3.3.1 基于剂量-体积准则的治疗计划的缺点.....	41
3.3.2 基于生物准则的治疗计划的优点.....	41
3.4 等效凸准则.....	42
3.4.1 等效凸物理准则.....	42
3.4.2 等效凸生物准则.....	42
3.5 目标函数.....	44
3.5.1 物理子目标函数.....	45

3.5.2 生物子目标函数.....	45
3.6 基于等效凸 NTCP 准则的 IMRT 方案优化算法研究	46
3.6.1 构造总目标函数.....	47
3.6.2 求解 FMO 问题.....	47
3.6.3 对比实验.....	48
3.6.4 实验材料及评价准则.....	49
3.6.5 实验结果.....	50
3.6.6 讨论.....	55
3.7 本章小结.....	57
3.8 参考文献.....	57
第四章 IMRT 中校正随机几何不确定性算法的研究.....	62
4.1 引言.....	62
4.2 IMRT 中的几何不确定性.....	63
4.3 利用反卷积算法校正 IMRT 中随机几何不确定性.....	64
4.3.1 1D 注量图反卷积算法.....	65
4.3.2 2D 注量图反卷积算法.....	66
4.4 改进的反卷积算法.....	66
4.4.1 算法描述.....	66
4.4.2 实验结果.....	69
4.4.3 讨论.....	75
4.5 本章小结.....	76
4.6 参考文献.....	76
第五章 自适应放射治疗中确定最佳重新规划时间及次数算法的研究.....	78
5.1 引言.....	78
5.2 自适应放射治疗技术.....	78
5.2.1 图像引导.....	79
5.2.2 计划调整.....	80
5.2.3 剂量验证.....	82
5.3 累加剂量.....	82
5.3.1 CT 数据配准.....	82
5.3.2 分次照射剂量分布校正.....	83
5.3.3 剂量累加.....	84
5.4 头颈部肿瘤自适应放射治疗中的最佳重新规划时间及次数.....	84
5.4.1 实验设计.....	84
5.4.2 实验结果.....	86
5.4.3 讨论.....	93
5.5 本章小结.....	94
5.6 参考文献.....	94
第六章 总结和展望.....	97
6.1 全文工作总结.....	97
6.1.1 改进的笔形束剂量计算方法.....	97
6.1.2 基于生物准则的 IMRT 方案优化算法.....	97
6.1.3 IMRT 中随机几何不确定性的校正.....	98
6.1.4 自适应放射治疗中的最佳重新规划时间及次数.....	98
6.2 未来工作展望.....	98
致 谢.....	100

个人简历.....	101
攻读博士学位期间发表论文清单.....	102

本文常用略语说明

AAA	Anisotropic analytical algorithm	各向异性分析算法
ART	Adaptive radiotherapy	自适应放射治疗
CT	Computed tomography	计算机断层成像
CTV	Clinical target volume	临床靶区
DAO	Direct aperture optimization	直接子野优化
DV	Dose-volume	剂量-体积
DVH	Dose volume histogram	剂量体积直方图
EUD	Equivalent uniform dose	等效均匀剂量
FFTC	Fast Fourier transform convolution	快速傅立叶变换
FMO	Fluence map optimization	注量图优化
IGRT	Image guided radiation therapy	图像引导放射治疗
IMRT	Intensity-modulated radiation therapy	强度调节放射治疗
	limited-memory	
L-BFGS	Broyden-Fletcher-Goldfarb-Shanno algorithm	L-BFGS 算法
MC	Monte Carlo	蒙特卡罗
MLC	Multi-leaf collimator	多叶准直器
NTCP	Normal tissue complication probability	正常组织并发症概率
PTV	Planning target volume	计划靶区
TCP	Tumor control probability	肿瘤控制概率

第一章 绪论

1.1 课题研究的背景和意义

恶性肿瘤已成为一种严重威胁人类健康且常见多发的疾病。国际肿瘤机构和世界卫生组织发表的统计数字显示,2008 年在全世界范围内新增了大约 1000 万的癌症患者,超过半数的癌症患者生活在发展中国家,且呈不断上升趋势,到 2015 年每年新增的癌症患者将增加到 1500 万,其中 2/3 的患者来自于发展中国家^[1]。我国的癌症防治任务也十分严峻。据相关统计资料显示,近 20 年来中国癌症死亡率上升了近 30%,2008 年全国癌症新发病病例达到 282 万,癌症死亡人数 196 万,位居全部死因的榜首;到 2030 年,每年的新发癌症病人将达到 2100 万,且每年的死亡人数将猛增到 1300 万^[2]。

目前,放射治疗已成为治疗肿瘤的三大主要手段之一。放射治疗主要利用射线穿透人体组织时,正常组织和肿瘤细胞对放射线所产生的不同生物反应机理来控制治疗肿瘤细胞的生长和增殖。电离辐射在人体受照射区域(靶区)内沉积能量,以伤害或破坏该区域内肿瘤细胞的基因物质,从而使其失去继续生长和分化繁殖的能力。与手术治疗和化学治疗相比,放射治疗可保留被治疗器官的功能,治疗副作用小,能够有效地保证患者愈后生存质量。据世界卫生组织统计,目前恶性肿瘤的治愈率约为 55%,其中放射治疗的贡献度达 40%^[3]。放射治疗的有效性使得其在肿瘤临床治疗中有广泛的应用。国内外统计数字表明,约有 50%-70% 的癌症患者需要接受不同程度(单纯放射治疗或与手术、药物配合治疗)的放射治疗^[4]。临床中使用的电离辐射可以是放射性元素所产生的 α 、 β 、 γ 射线,也可以是由 X 射线治疗机、医用加速器产生的 X 射线、电子线、质子束、中子束、介子束,甚至其它重带电粒子束。本文主要研究外照射光子束(X 射线)对肿瘤细胞的放射治疗。

放射治疗的根本目的是努力提高放射治疗的治疗增益比。给予病灶(靶区)足够大和足够精确的照射剂量,杀灭肿瘤细胞,同时使靶区周围的正常组织(normal tissue)和危及器官(organ-at-risks, OARs)免受或尽可能少地接收到不必要的射线照射,即在提高肿瘤控制率(tumor control probability, TCP)的同时,减少周围正常组织出现并发症的概率(normal tissue complication probability, NTCP)。近年来,随着计算机技术、医学影像技术、图像处理技术和加速器相关硬件制造技术的飞速发展,在常规放射治疗的基础上,出现了许多放射治疗新技术、新手段,如立体定向放射外科(主要包括 X 刀、 γ 刀)、三维适形放射治疗(3-Dimension conformal radiation therapy, 3D-CRT)、调强放射治疗(intensity modulated radiation therapy, IMRT)、图像引导调强放射治疗(image-guided radiation therapy, IGRT)和剂量引导调强放射治疗(dose-guided radiation therapy, DGRT)等。放射治疗逐步朝着“精确定位,精确计划,精确治疗”的方向发展^[5]。放射治疗的精度最终取决于受辐射部位的剂量分布,准确地预测剂量分布是放射治疗质量保证的关键环节。在现代放射治疗中,这一重要任务是由放射治疗计划系统(radiotherapy treatment planning system, TPS)来承担的。放射治疗计划系统就是在高性能计算机上制定放射治疗方案的软件系统。通过对诊断和治疗信息进行整合,放射治疗系统模拟整个治疗过程,制定出最佳的放射治疗方案,以保证靶区内肿瘤细胞在得到足够致死剂量的同时,周围正常组织接收尽可能少的受照射量。一个完整的、高性能的放射治疗系统需要大量关键技术的支持,如医学图像的快速三维重建、高精度的三维剂量计算、治疗计划的方案优化或逆向设计、功能强大的放射治疗设备等。放射治疗计划系统的发展水平在一定程度上代表着临床放射治疗技术的发展水平。本文主要针对剂量计算、IMRT 方案优化、IMRT 中几何不确定性的校正和自适应放射治疗四个方面展开了研究。

1.2 国内外研究现状

1.2.1 光子剂量计算方法

在放射治疗中,吸收剂量是度量单位质量受照物质吸收辐射能量多少的一个量。由辐射作用于物质所引起的效应主要取决于该物质所吸收的辐射能量,因此吸收剂量在辐射效应的研究中极为重要。精确快速的 3D 剂量计算是现代放射治疗计划系统的核心部分。剂量计算是连接放射治疗设备所选择参数和实际临床治疗效果的唯一纽带。剂量算法在实际应用中主要有两方面:首先是将剂量算法应用到方案优化中,得到治疗设备所需的参数;然后,根据方案优化得到的设备参数再次计算剂量分布,并根据剂量分布来评价所选择的设备参数是否合适。剂量计算精度直接影响到放射物理师所制定治疗计划的质量,以致最终影响放疗的精度与疗效。考虑到临床上的实际需要,计划系统对剂量计算的精度和速度有着严格的要求。放射治疗后,约有近半数患者在肿瘤没有扩散的情况下仍然死亡,放射治疗中剂量分布的不准确是病人受照后死亡的一个主要原因^[6]。Boyer 和 Schultheiss 的研究结果表明:照射剂量的准确性提高 1%,肿瘤的治愈率便可提高 2%^[7]。一般把 ICRU (international commission radiation units & measurements) 24 号报告中推荐的 $\pm 5\%$ 作为照射剂量误差的允许范围^[8]。在逆向计划中,优化过程需要进行多次的剂量计算(大约 10 到 1000 次)^[9],因此对计算速度的要求也非常苛刻。一个具备临床实用性的剂量计算模型应在 1 分钟以内完成单野、低精度的剂量计算;在 1 小时内完成多野、高精度或优化剂量计算^[6]。

随着计算机存储容量和运算速度的不断提高,剂量计算方法也在逐步朝着更高精度的方向发展。从早期基于修正(correction-based)的剂量计算模式向基于模型的半解析算法转变,最终朝着直接解析算法发展^[10]。上世纪 50 年代,高能光子束刚被引入医学领域的时候,剂量计算均采用经验方法,主要包括矩阵法和函数法。矩阵法根据参考条件下的大量实验测量数据对病人体内的剂量分布进行计算。但是由于矩阵法需要存储大量实验数据,对计算机硬件有着较高的要求,为此,人们又提出了函数法。函数法使用数学解析式直接计算某一点的相对剂量,不需要存储大量数据。由于以上两种算法存储的数据依赖于实验条件和实验模型,在用于临床剂量计算时通常需要根据实际情况进行校正,所以这两种算法也被称为基于修正的剂量计算方法。基于修正的剂量计算方法由于其概念简单,而且计算结果较为可靠,目前仍然是放射治疗计划系统中应用广泛的剂量计算方法。半解析算法的基本原理是通过等效路径(equivalent path length, EPL)对散射粒子输送问题进行间接模拟。由于半解析算法没有对散射粒子进行完全精确地模拟,所需的计算量也比较小。现在临床上常用的剂量算法就是半解析算法中基于核模型的算法。但是,这类算法不能完全精确地计算非均匀介质中的剂量分布,特别是在密度变化较大的区域内^[11]。在研究半解析算法的同时,学者们也在研究另一种更为直观的方法:直接解析方法,简称为解析法。这类方法对散射粒子传输进行了精确模拟,从而可以得到更为精确的剂量分布。解析方法可分为两类:随机模型(蒙特卡罗方法)和粒子输运方程。

1.2.1.1 半解析算法

半解析算法通常可分为三类:差分散空比模型、德尔塔模型和基于核的模型。前两种方法都有各自的局限性,并没有得到广泛的研究。现代医疗系统中常用的是基于核模型的方法。基于核模型的卷积/叠加剂量计算的概念来自于图像处理领域。一束射线进入体模之后,其能量会随着射线入射深度的增加而减少。部分进入体模的射线与体模内的物质反生碰撞,而其余的继续沿射线方向前进。基于核模型的方法认为发生碰撞的射线的能量会沉积在碰撞点周围,能量沉积核(energy deposition kernels, EDK)就是用来描述这部分能量在碰撞点周围沉积的概率密度函数。类似于图像处理中卷积/叠加的概念,当前点的剂量沉积被认为是对周围碰撞点的反映(能量沉积核)的适当加权之和。根据不同的能量沉积模式,能量沉积核可分为三类:点核(point kernel),笔形束核(pencil beam kernel)和面核(planar kernel)。

1.2.1.1.1 点核模型

点核是描述原射线发生碰撞后,碰撞释放的能量在碰撞点周围无限大介质中的能量沉积模式。Mackie 等将其称为剂量扩散矩阵(dose spread array)^[12]。用蒙特卡罗(Monte Carlo, MC)方法模拟得到了原射线在水体模内发生碰撞后,碰撞释放的能量在碰撞点周围扩散沉积的能量沉积矩阵;将这个沉积矩阵与体模中每点处原射线释放的碰撞能量卷积得到整体的剂量分布^[12]。Mohan 等称其为微分笔形束核(differential pencil Beam)^[13]。Ahnesjö 等将其称为点扩散函数(point spread function)^[14]。在商业治疗计划系统中(如 Nucletron Helax-TMS 和 Philips Pinnacle TPS),广泛使用的基于点核模型的剂量计算方法是 Ahnesjö 等在 1989 年提出来的利用筒串卷积技术(collapsed cone)计算剂量分布的方法^[15]。

利用点核进行剂量分布计算的运算量很大。在均匀组织中计算 N^3 个点的剂量需要 N^6 次乘法和加法运算,如非均匀组织中计算则需要 N^7 次操作。大量的计算限制了点核方法在临床上的应用。为此一些学者提出了用快速算法加速剂量计算,如:Boyer和Mok等提出了用空间不变点核计算非均匀组织内剂量分布的方法,这样可用快速傅立叶变换进行离散卷积,计算量减少到 $N^3 \log_2 N$ ^[16]。Erdi等用3D Hartley变换加速剂量计算,计算量节省了1000多倍^[17]。但是这类使用快速变换的方法不能精确计算原射线部分的剂量分布,因为原射线核有较陡峭的梯度变化。Ahnesjö提出的筒串卷积技术在保证计算精度的同时减少了计算量^[15]。筒串卷积技术将点核划分成具有一定立体角的共轴锥筒串,碰撞释放的能量沿筒轴方向线性传输,并被筒轴上的体积单元衰减和吸收。Lu等人从主要从两方面改进了筒串卷积方法的精度:使用CCK(cumulative-cumulative kernel)计算剂量分布,提高剂量计算精度;在剂量叠加过程中考虑到了体素大小对剂量计算精度的影响^[18]。随着科技的发展,一些学者尝试利用硬件加速的方法加快剂量计算速度。硬件加速的方法主要是对筒串卷积法和蒙特卡罗法进行加速。对筒串卷积法进行加速的方法有:Whitton和Zhou等人用可编程门阵列(field programmable gate array, FPGA)加速剂量计算^[19, 20];de Greef和Hissoiny等人用图形处理器(graphics processing unit, GPU)加速剂量计算^[21-23]。

1.2.1.1.2 笔形束核模型

由于基于点核模型的3D剂量计算所需计算时间过长,一些学者提出了基于笔形束核的2D剂量计算方法。笔形束核描述的是一束射线入射在无限大均匀介质中的能量扩散模式,笔形束核可以看成对点核在入射方向上进行的卷积。由于基于笔形束核的剂量计算是二维卷积,相比基于点核模型的三维卷积剂量计算,运算量大为减少,因此被广泛用于方案优化算法中。

笔形束核通常是利用蒙特卡罗方法模拟得到的,直接模拟得到或将点核在入射方向进行卷积得到^[12-14, 24]。由于这种方法是精确地模拟光子和电子在体模内的物理输运过程得到的,因此所得到的笔形束核精度比较高。但是,蒙特卡罗方法模拟得到的单能笔形束核不能直接在临床中应用,因为由实际射线源发出的射线是多能谱射线;而且射线入射到体模表面,在体模表面的注量分布与离轴距离(off-axis)是有关的;同时还要考虑射线在空气中的电离污染等。在剂量计算过程中,剂量计算的精度很大程度上取决于对治疗头的精确模拟。Liu等人将射线源模拟为两个源:原光子射线点源表示直接来自于靶目标的光子射线和虚光子射线点源表示来自于初级准直器和均整滤波器的散射光子射线^[25]。Yang等将实际射线源模拟为3个点源:一个原光子射线点源表示来自于靶目标的原光子射线,一个虚光子射线点源表示来自于初级准直器的散射光子射线和一个虚光子射线点源表示来自于均整滤波器的散射光子射线^[26]。Lillikainen等人也将射线源模拟成3个点源:一个原光子射线点源表示在靶目标上韧致辐射产生的原光子射线,一个虚光子射线点源表示来自于加速器头、均整滤波器、初级准直器和次级准直器的散射光子射线和一个虚电子射线点源表示电离污染^[27]。

鉴于直接利用蒙特卡罗方法模拟笔形束核的诸多困难,一些学者直接从测量数据中获取笔形束核^[28-31]。但是,在小射野情况下,侧向电离平衡没有达到,这种方法的精度不高。Chui等人用反卷积的方法从测量得到的宽束射线剂量分布中提取笔形束核^[28]。为了避免最后

得到笔形束核数据的波动,对射束数据进行了预处理,使得其关于转折点对称。Bortfeld 通过微分输出因子和组织体模比(tissue phantom ratio, TPR)的乘积来得到笔形束核^[29]。该方法认为组织体模比是一个关于深度方向变化的函数,且不受射野大小的影响,并使用输出因子来校正小射野情况下的电离不平衡。Ceberg 等人在圆形射野下通过微分散射原射线比(scatter-to-primary ratio, SPR)得到笔形束核,笔形束核表示为深度和离轴距离的函数^[30]。因为散射原射线比是在电离达到平衡的条件下测量得到的,因此 Ceberg 等人将一个电离平衡因子加到散射原射线比上来校正电离不平衡区域的剂量分布。Storchi 提出一种新的利用测量数据得到笔形束核的方法,该方法所需的测量数据包括:中心轴处的深度剂量信息、体模散射比和离轴比等^[31],与 Bortfeld 和 Ceberg 的方法相比,在小射野电离不平衡的情况下,笔形束核是根据半影区域的离轴比,而不再是中心轴剂量信息计算得到。

在调强放射治疗中需要使用特定形状、位置的多叶准直器来构成调强放射治疗的射野。而后者可以被认为是由有限大小的注量单元组成。Bourland 和 Chaney 等人对此进行了研究,他们认为由放射源发出的射线束可以被分为多个“有限笔形束”,而射线束在体模内产生的总的剂量沉积可以通过对多个“有限笔形束”在体模内产生的剂量沉积叠加获得^[32]。Ostapiak 等人改进了这类方法,改进后的方法使用快速傅立叶变换技术来加快剂量计算速度;同时使用多能谱射线源,并考虑到了射线源的大小和粒子穿过多叶准直器时的散射对剂量计算的影响,从而提高了剂量计算精度^[33]。这些“有限笔形束”可以构成任意形状的射野,而不再依赖于计算笔形束核时所使用的射野大小。但这类方法仍不能满足调强放射治疗计划对剂量计算速度和半影区处理的要求。Jeleń 等人提出了一种新的有限笔形束(finite size pencil beam, fsPB)算法,该算法将笔形束核的深度信息存储在查找表中,并用解析函数模拟笔形束核的侧向剂量分布,从而可以增加半影区剂量计算的精度和加快剂量计算的速度^[34]。随后, Jeleń 等人又对该算法继续进行了改进,通过 3D 密度校正的方法有效地提高了原有算法在非均匀介质中的剂量计算精度^[35]。

各向异性分析算法(anisotropic analytical algorithm, AAA)是一种新型三维笔形束卷积/叠加算法,最早由 Ulmer 等人提出数学模型^[36],随后经瓦里安医疗系统小组继续发展,并在 Eclipse(Varian medical system, Palo Alto, CA)放射治疗系统中使用。相比传统的笔形束算法,AAA 算法考虑到了治疗头发射出的射线在到达体模前的散射和电子污染,以及光子在非均匀介质中的各向异性分布,从而提高了剂量计算的精度。AAA 算法主要由两部分组成:射线源的模拟和体模内的剂量计算。Lillikainen 等人对 AAA 算法进行了改进,使用六个径向指数函数组合形式对侧向散射核进行模拟,取代原算法中使用的三个高斯函数组合形式^[37]。改进后的算法降低了在组织边界处产生的误差,提高了剂量计算的精度。同时有很多文献对 AAA 算法在不同体模内的剂量计算精度进行了报道^[38-42]。研究结果表明:在非均匀介质中,AAA 算法的剂量计算精度要明显优于传统笔形束算法,特别是在小照射野、高能量射束、射束的半影区以及有低密度非均匀物质存在的情况下。但 AAA 算法的计算模型是基于假设深度方向和侧向方向相互独立的笔形束模型,不能精确地模拟粒子在体模内的输运过程,从而不能准确计算非均匀介质中的剂量分布。

基于笔形束模型的剂量计算方法可以很好地处理不规则射野和射野适形等问题下的剂量计算,但在考虑人体非均匀组织和患者体表不规则轮廓的情况下,仍需要使用宽束校正方法对其进行校正,以此提高剂量计算精度。Li 等人认为非均匀介质中剂量分布与所使用的射束能量、射野大小、介质密度等因素呈非线性关系,因此密度拉伸法并不能完全校正非均匀介质中的剂量分布^[43]。AAPM 85 号报告对剂量计算方法中不同的非均匀介质校正方法进行了总结和分类^[44]。第一类是基于局部能量沉积的方法,主要包括线性衰减法、RTAR(ratio of TAR (tissue air ratio))法、Batho 幂法、ETAR(equivalent TAR)法、DSAR(differential scatter air ratio)法和 DVOL(delta volume)法等。这类方法主要根据介质密度对原射线注量进行校正,因此对散射部分只能做处理近似。这类方法考虑了粒子在体模内传输的能量衰减变化,但没有考虑次级电子的传输情况,因此在处理高能量射线剂量分布时误差较大。第二类是基于非局部能

量沉积的方法，主要包括卷积叠加法、快速变换法和 dTAR(differential TAR)法等。这类方法同时考虑了注量和剂量扩展函数的变化，并可以通过对核函数的校正来获得更加精确的剂量计算结果。

1.2.1.2 直接解析算法

1.2.1.2.1 蒙特卡罗算法

蒙特卡罗算法，又称随机抽样技术或统计试验方法，是通过使用随机数来解决数学和物理问题，其中抽样的大小将决定统计的误差，因此也影响结果的准确性。粒子与物质相互作用时服从统计学规律，如发生作用的位置、作用的形式、发生作用后粒子可能被吸收或散射、散射粒子的运动方向和能量、两次作用位置间的距离等参数均是随机变量^[4]。因此，可以用蒙特卡罗方法来模拟粒子与物质相互作用的全过程。通过模拟大量(10 万甚至 100 万个)粒子的运输过程，就可以精确地计算出粒子束与物质相互作用的宏观特征，如注量(fluence)分布、吸收剂量分布。目前蒙特卡罗方法及程序已在医学剂量计算及验证方面得到广泛应用，并被公认为是当前所有剂量计算方法中最精确的一种，在医学领域被称为剂量计算的“金标准”(gold standard)。蒙特卡罗算法的优点是可以处理粒子运输的各种复杂情况，尤其是一些难以进行实验测量的情况。目前在肿瘤放射物理学中的应用主要有：1) 外照射射线源模拟；2) 剂量仪响应模拟；3) 外照射时，体模内辐射场的模拟；4) 外照射治疗计划应用，如验证算法、提供配置数据（如计算笔形束的剂量分布）等；5) 腔内放射治疗源周围辐射场的模拟。

目前广泛应用于放射治疗剂量的蒙特卡罗程序有，由美国国家研究院 Los Alamos 所开发的 MCNP(Monte Carlo N-Particle transport code)^[45]、由欧洲共同体核能研究机构(CERN)与日本高能加速器研究机构(KEK)主导，20 多家机构共同开发的所开发的 GEANT4(geometry and tracking version 4)^[46]、意大利核物理研究机构(INFN)和欧洲核子研究中心(CERN)研制开发的 FLUKA(FLUktuierende KAskade)^[47]、由美国斯坦福大学直线加速器中心(SLAC)所开发的 EGS4(electron gamma shower version 4)^[48]、由加拿大国家研究院(NRCC)与美国威斯康辛大学合作的 OMEGA 计划所开发的以 EGS4 为基础开发的放射治疗剂量蒙特卡罗模拟软件 OMEGA/BEAM^[49]及由美国放射安全信息中心(radiation safety information computational center, RSICC)开发的 PENELOPE 蒙特卡罗剂量计算软件^[50]等。但是蒙特卡罗方法存在收敛速度慢、计算时间长的弱点，阻碍了其在临床治疗计划中的广泛应用。由于蒙特卡罗方法非常适合进行并行计算，一些学者利用图形处理器(graphics processing unit, GPU)加速蒙特卡罗方法^[51-54]。与普通 CPU(central processing unit)相比，图形处理器能提高剂量计算速度大约 400-800 倍，有些算法单野剂量计算甚至能在 1 分钟内完成^[54]。

1.2.1.2.2 粒子输运方程

另一种直接解析方法是解粒子输运方程，即玻尔兹曼输运方程(boltzmann transport equation, BTE)。玻尔兹曼输运方程是对粒子与物质之间的相互作用过程进行精确模拟，可以精确地模拟粒子在非均匀体模中的传输和散射作用。由于高能粒子在体模内以光速通过，与体模内物质的相互作用的概率很小且可以被忽略，因此可以使用线性玻尔兹曼输运方程计算体模内或患者体内的剂量分布。在实际应用中需要考虑脉冲放射源、核反应动力学等因素，此时玻尔兹曼输运方程还是关于时间的函数。一些学者对玻耳兹曼输运方程的求解进行了研究，并在放射治疗领域实现了可供使用的程序^[55-58]。2010年Varian公司已将解粒子输运方程的剂量算法(Acuros[®] XB)应用于放射治疗系统并在临床治疗中使用。一些学者将这类算法与常用的基于点核和笔形束的剂量算法及蒙特卡罗方法进行了比较^[59, 60]。解粒子输运方程的方法能达到和蒙特卡罗方法一样的剂量精度。值得注意的是，蒙特卡罗算法和解粒子输运方程的算法都是基于玻尔兹曼输运方程的算法。蒙特卡罗算法是模拟大量样本来代替真实的结果，而解粒子输运方程的方法顾名思义就是直接解玻尔兹曼输运方程。相比蒙特卡罗算法，解粒子输运方程的剂量计算方法更加有效^[57]，尤其是在求解逆问题的时候，利用解粒子输运方程

的方法可以直接求解逆问题；而蒙特卡罗算法则需要多次迭代来求解逆问题。因此，基于粒子输运方程的剂量计算方法在未来具有很大的研究价值。

1.2.2 IMRT 方案优化算法

一个较好的治疗计划应该满足临床剂量学四原则^[4]：1)肿瘤剂量要求准确；2)肿瘤治疗区域内的剂量分布要均匀，剂量变化不能超过 $\pm 5\%$ ，即要达到 90% 的剂量分布；3)照射野设计，应尽量提高治疗区域内剂量，降低照射区正常组织的受剂量范围；4)保护肿瘤周围重要器官（如食道癌治疗时保护脊髓）免受照射，至少不能使他们接受超过其允许耐受剂量范围的照射。早期的治疗计划是采用正向设计、手工规划的方法，医生或计划设计者根据治疗方案的要求，利用自己的经验选择射线类型及能量、射野个数、形状及入射方向、射野剂量权重、外加射野挡块或楔形板等参数，由计算机计算体模内的剂量分布，然后对计划进行评估，这是一个反复试误(trial-and-error)的过程，在找到合适的方案之前要一直进行下去，这种方法非常耗时，而且很难达到满意的效果。而逆向计划则正好相反，它是根据医生确定的治疗要求，由计算机通过各种优化算法来自动寻找出符合要求的最佳治疗方案。方案优化是求解受限目标函数最大或最小值的过程，找到目标函数达到最大或最小值时所对应自变量的值。

在放射治疗的逆向计划当中，约束函数一般由物理准则(physical criteria)或生物准则(biological criteria)构成。物理约束函数通常定义为处方剂量分布与所计算剂量分布之差的平方，是通过给定或限定靶区和危及器官应达到的物理剂量分布，实施准确的治疗。常用的物理准则有最小剂量约束函数、平均剂量约束函数、最大剂量约束函数和剂量-体积(dose-volume, DV) 约束函数。物理约束函数的优点是定义清晰简单，易于在临床上应用。然而物理约束函数不能准确地预测肿瘤或正常组织内的非线性放射生物效应，特别是在非均匀剂量照射的情况下。而生物约束函数则是由生物准则构成，是通过给定或限定肿瘤控制率和正常组织并发症发生的概率，找到控制肿瘤和减少正常组织并发症之间最好的折衷，实施最佳的治疗。生物约束函数能够预测肿瘤或正常组织对实际剂量分布的放射生物效应，但生物准则中所使用的参数需要进行提前校正，否则计算得到的结果将失去生物意义^[61]。常用的生物准则有等效均匀剂量((generalized) equivalent uniform dose, (g)EUD)、正常组织并发症概率(NTCP)和肿瘤控制概率(TCP)以及其它加权准则等^[62]。

IMRT 的主要实现方式包括：二维物理补偿器调强、多叶光栅(multileaf collimator, MLC)调强、断层调强放疗、电磁扫描调强放疗、二维调强准直器、独立准直器静态调强等^[2]。当前临床中使用最为广泛的是 MLC 调强技术。MLC 调强技术主要有 MLC 静态调强和 MLC 动态调强两种模式。静态调强是把射野强度分布分割成一系列的子野，利用 MLC 对形成的多个子野进行分步照射，其特征是每个子野照射完毕后，切断照射，MLC 调到另一个子野位置，再继续照射，直到所有子野照射完毕。所有子野强度分布的累加，形成要求的强度分布^[4, 63]。静态调强有硬件实施容易控制，可靠性高等优点^[64, 65]，但其不足之处在于，对于多峰值的射野强度分布（大多数情况），所需子野个数多，延长了总治疗时间(包括照射时间和中间的过渡时间)，射线利用效率低。动态调强则利用多叶光栅相对应的一对叶片的相对运动，实现对射野强度的调节，其特征是叶片运动过程中，射线一直处于“BEAM ON”的位置，剂量率不变。动态调强方法包括：动态叶片(dynamic-leaf collimation)、调强旋转(volumetric modulated arc therapy, VMAT)、动态 MLC 扫描(scanning-leaf)等方法。动态调强的优点在于可以实现连续的强度调节，其不足在于对设备性能要求高、质量监控困难等。

传统的静态调强实现方法主要有：两步法和直接子野优化法(direct aperture optimization, DAO)两种。两步法，顾名思义，主要是分两步来实现的：1) 注量图优化(fluence map optimization, FMO)，使用优化算法求得最优注量图分布；2) 注量图分割，将最优注量图分割成一系列可被 MLC 执行的子野。两步方法的一个主要优点是 FMO 可以将被优化问题归结为一个连续的凸优化问题，并可以使用一个高效的算法来解决优化问题。在注量图优化过程中，一些准则(生物准则)是关于注量分布的非凸函数。为了能在 FMO 中使用这些非凸准则并且能

使用高效的基于梯度的优化算法解决 FMO 问题, 一些学者首先将这些非凸准则变换成等效凸准则, 然后应用到 FMO 中^[66-71]。Wu 等人将剂量-体积准则转换为等效的凸物理准则^[66], 这种方法已广泛地应用于临床放射治疗中。随后, Wu 等人证明这种转换方法会导致优化过程中存在多个局部最优点, 但是这并不影响在实际临床治疗中使用基于梯度的优化算法找到全局最优点^[67]。Wu 等人通过数学变换将 EUD 准则变换为等效凸准则^[68]。Choi 等人证明了, 当参数 $a \geq 1$ 时, EUD 准则为凸函数; 反之为凹函数^[69]。Romeijn 等人^[70]和 Hoffmann 等人^[71]分别将不同类型的生物准则转换为对应的等效凸生物准则。在静态模式中, 过多的子野个数将延长治疗时间, 过多的总机器跳数将增加叶片间的射线泄露。因此, 如何在保证投射剂量的前提下减少子野个数和总机器跳数是注量图分割的两大目标^[72]。与此同时, 为了减少叶片间的凸凹槽效应和提高射线的利用率, 在满足上述目标的基础上, 子野的开口面积也应该尽可能的大。Baatar 等人已证明将注量图分割为最少个数的子野是一个 NP 困难(non-deterministic polynomial-time hard, NP-hard)问题^[73]。已有算法主要是用自适应算法^[74]和规划类算法^[75]来解决这类问题。两步法的主要缺点在于: FMO 过程中射野是由独立的射束组成和注量图分割过程, 这两个过程都存在着降低治疗质量的危险。因此, 一些学者将基于射束的 FMO 和注量图分割整合在一个优化模型中^[76, 77], 这种方法被称为 DAO。与两步法相比, DAO 方法在保证治疗质量同时, 有效地减少了放射治疗所需要的子野个数, DAO 方法可以直接得到每个子野的形状和对应的机器跳数值, 从而减少了总的治疗时间。但 DAO 方法增加了计算量, 延长了优化过程所需的时间。现在常用的 DAO 方法主要有 3 种:

1) 随机搜索方法: 随机改变准直器叶片的位置。如果改变后的射野形状改进了目标函数的值, 那么就接受当前的更改; 反之, 为了能跳过局部最优点, 就按一定的随机概率拒接当前的改变。这种方法最早是由 Shepard 等人^[78]提出并使用模拟退火方法解决这类优化问题, 随后 Earl 等人^[79]对其进行了改进。Li 等人^[80]和 Cotrutz 等人^[81]分别用遗传算法解决这类优化问题。

2) 基于局部梯度的方法: 将 MLC 叶片位置作为优化的自变量。这种方法需要建立目标函数和 MLC 叶片位置之间的关系并给出其一阶导数。这种方法已经在一些商用治疗系统中使用, 如 Pinnacle 和 Ray Station 中使用的直接机器参数优化(direct machine parameter optimization, DMPO)模型^[82]。这类方法的难点在于如何给出目标函数和 MLC 叶片位置之间的关系, 到目前为止还没有文献对此关系进行详细描述。

3) Column-generation 法: 这种方法在迭代开始的时候没有设置任何初始的子野, 随着迭代的进行, 子野逐个被加入到治疗方案中。这种方法在每次迭代过程中分两步来实现: 解代价问题得到在当前情况下最优的子野形状并将该子野加入到治疗方案中, 最优的子野是能够最大程度地减小目标函数的值, 且同时能被 MLC 执行的射野形状子野; 解主问题得到方案中每个子野的权重值。已有一些学者对这种方法进行了研究^[76, 83-85], 而且 Men 等人利用 GPU 技术对该方法进行了加速^[86]。

1.2.3 IMRT 中几何不确定性的校正

放射治疗过程中的几何不确定性主要源于三个方面: 摆位误差、器官运动和形变引入的误差、机械误差。几何不确定性可被分为系统不确定性和随机不确定性两类^[87]。如果每次分次放射治疗前均使用在线 IGRT 技术, 既可以减小系统摆位误差, 也可以减小随机摆位误差。自适应放射治疗(adaptive radiotherapy, ART)可以在一定程度上减小系统摆位误差, 但对随机摆位误差的影响较小。实时跟踪放射治疗技术(real-time motion tracking)主要针对分次放射治疗内靶区中运动幅度较大的肿块进行校正, 比如肺癌和前列腺癌。由于该技术需要一定的反应时间和使用预先设计好的放射治疗计划, 因此也很难完全地校正随机不确定性。因此, 放射治疗中的随机不确定性的来源一般包括: 如果分次放射治疗前没有使用在线 IGRT 技术, 那么摆位误差里面包含了部分随机误差; 分次放射治疗内患者靶区的不规律运动和随机的形状变化也是随机误差来源; 放射治疗使用的医用直线加速器的机械精度需要定期进行校正,

校正后的机械误差一般以随机误差为主。遵循自适应放射治疗准则和在日常摆位中使用 IGRT 技术可以明显地减少系统不确定性^[88, 89]。但使用 IGRT 技术和实时跟踪放射治疗技术却只能减少部分随机不确定性^[90-92]，这主要是由于放射治疗设备反应灵敏度差以及随机不确定性的运动幅度和形式造成的。

校正随机不确定性的传统方法是将临床靶区(clinical target volume, CTV)扩展到计划靶区(planning target volume, PTV)来覆盖所有靶目标可能移动到的位置。这种方法容易使靶目标周围正常组织受到高剂量照射，尤其是扩展部分覆盖了部分 OARs，增加了并发症发生的概率，从而降低了放射治疗的质量。如果不扩展 CTV，靶区边缘处的组织在移动到射野之外时，接收到的低剂量照射，又会降低肿瘤控制率。随机不确定性被认为近似地服从高斯分布^[93]。一些学者利用随机不确定性的概率密度分布形式改进了这种扩展靶区的方法，使靶区扩展部分接收到的剂量随着到靶目标距离的增加而减少，如随机模拟法^[94]、剂量卷积法^[95, 96]和注量图修改法^[97]。但是，改进后的方法仍不可避免地扩大了 CTV 的覆盖范围，于是另一些学者使用鲁棒注量图的方法明显地减少了 CTV 扩展部分的覆盖范围^[98-103]。鲁棒注量图法是在射野边缘处给与高剂量照射。因此，靶区边缘处的组织在移动进入射野内时接收到的高剂量可以补偿其移动离开靶区时所接收到的低剂量。其整体的输出结果是在不扩展 CTV 的情况下，靶区在考虑随机不确定性情况下接收到的剂量，与靶区在静止不动时应该接收到的处方剂量相同。实现鲁棒注量图的方法主要有解析函数法^[98, 99]、优化方法^[100-107]和序列反卷积^[108]三类。Lind 等人通过简单的解析函数调整每个射野方向的注量分布，使得在存在随机误差的情况下靶目标仍然能接收到处方剂量分布^[98, 99]，实验结果表明这种增加射野边缘剂量的方法可以有效地取代扩展 CTV 的方法，并能减少靶区周围正常组织接收到的剂量。利用随机不确定性的概率分布函数，Li 等人使用实际照射剂量的均值构造优化模型的目标函数^[100]，而 Unkelbach 等人使用实际照射剂量和处方照射剂量的差值构造优化模型的目标函数^[101, 102]。Chu 等人直接将靶区剂量接收到的概率分布引入到优化模型中并在前列腺肿瘤病例上验证了该算法的有效性^[103]。Baum 等人通过靶区的覆盖率将随机不确定性信息引入到优化模型中^[104]。Olafsson 等人将随机不确定性信息引入目标函数并使优化问题可以用线性规划的方法来求解优化^[105]。Chan 等人 and Bortfeld 等人分别用优化的方法求解鲁棒的注量分布，其中呼吸运动被认为是一种服从概率分布的几何不确定性^[106, 107]。Fan 等人使用序列反卷积的方法计算鲁棒的注量分布^[108]。

1.2.4 自适应放射治疗

IMRT 在放射治疗中的地位非常重要。IMRT 可以增加肿瘤剂量的覆盖并减少周围正常组织接收到的剂量，从而提高了放射治疗病人的治愈率^[109]。但是，病人在长达 6-7 周的治疗过程中，由于内部对放射治疗的反应以及自身内部器官的变化，使得靶目标和正常组织的位置、形状以及体积都发生了明显的变化。在高度调强的 IMRT 治疗方案中，这些变化很容易导致正常组织并发症的发生和肿瘤控制率的降低。因此，在放射治疗过程中，由于病人内部解剖结构的变化，在最初设计的 CT(computed tomography)数据上设计的治疗方案不再是最优的方案。Yan 等人于 1997 年在放射治疗过程中首次提出 ART 的概念^[110]。ART 是通过图像数据、剂量以及其它信号的反馈对治疗方案进行修正，从而对治疗过程中出现的误差进行校正^[111]。ART 的实现方式包括：在线移位、离线 ART、在线 ART 和混合算法。影像引导放射治疗(image-guided radiation therapy, IGRT)技术首先被用来校正日常放射治疗过程中由摆位及病人解剖结构变化所引起的误差。IGRT 已经在临床放射治疗中广泛使用。通过获取患者由不同方式(如锥束 CT、核磁共振图像等)得到的 3D 影像，可以精确地捕捉到患者体内解剖结构的变化。IGRT 通过移动剂量“云”的位置到计划靶区来保证靶区剂量的覆盖和避免正常组织受到高剂量的照射。但是，仅仅依靠病人位置的移动不能完全地校正体模内靶区和正常组织的变化，需要更复杂的操作来进行校正。

离线 ART 是在分次放射治疗后获取病人的 CT 数据，并在获取的 CT 数据上重新设计治疗方案，按新的方案实施后续分次照射，从而有效地校正系统移位误差和射线分次照射过程

中产生的剂量偏差^[112-116]。在离线放射治疗中, 总的剂量分布由 3 部分组成: 已照射的剂量、当前要照射的剂量和将来要照射的剂量。根据分次照射后立即获取的 CT 数据可以得到已照射剂量, 即已经实施分次照射的累积剂量。累积剂量与已经应该被照射的剂量之差就是要校正的剂量误差。这个剂量误差可以通过当前的一次分次照射或多次分次照射来校正。由于不能准确获取病人在将来的 CT 数据, 因此一般认为当前的分次照射和将来的分次照射不会引入系统误差, 即假设病人将来获取的 CT 数据与最近获取的 CT 数据是相同的。Wu 等人^[112]和 de la Zerda 等人^[114]分别给出了类似的基于物理准则的 ART 优化模型。de la Zerda 等人引入了反馈的概念并详细介绍了不同的离线 ART 模型^[114], 这些模型之间的主要区别在于所使用的 CT 数据以及校正系统误差所需要的分次照射次数不同。Sir 等人将生物准则(线性 EUD 模型)应用到了离线 ART 模型中^[115]。由于离线 ART 所使用的 CT 数据是在上次分次照射后得到的, 因此将离线 ART 方案应用到下次分次照射可能会引入一些误差。Yan 等人用最近 5 次得到的锥束(cone beam, CB)CT 数据来估算病人当前或将来的 CT 数据^[116]。在线 ART 在分次放射治疗前获取病人的 CT 数据并立即在获取的 CT 数据上调整治疗方案, 并按修改后的方案实施当前分次计划^[117-120]。Wu 等人根据初始 CT 数据和分次照射时得到 CT 数据中的靶目标和正常组织的轮廓在 beam's-eye-view (BEV)方向投影的差异, 得到在该投影方向二者之间的形变矩阵, 然后根据该形变矩阵校正初始方案在这个方向的子野形状, 并根据校正后的子野形状进行当前的分次照射^[117]。Court 等人通过靶目标和正常组织的刚性形变依次调整 MLC 每个叶片的位置^[118]。Feng 等人直接计算 3D 形变矩阵, 根据形变矩阵在各个射野方向的投影校正初始方案在该方向的子野形状^[119]。Ahunbay 等人在实时校正射野形状后, 还重新对所有子野的权重进行了优化^[120]。该方法在校正靶目标和正常组织在各射野投影方向的形变的同时, 还能校正射野深度方向的形变。实时 ART 的优点是能根据实时获取的病人数据修改方案从而保证最优的治疗, 然而其主要技术难题是要求治疗方案的修改要求在几分钟之内完成, 因为病人正躺在治疗床上等待治疗。离线 ART 和在线 ART 方法大大地增加了物理师和医生的工作量, 限制了其在临床实践中的使用。在刚性靶目标变化或器官形变较小的情况下, 在线移位仍然是一种非常有效的校正方法。因此, 将在线移位方法和离线/在线 ART 方法组合成的混合方法, 可以在尽可能提高计划质量的同时减少重新规划所需冗长的时间^[121, 122]。

1.3 本文工作内容

本论文主要围绕着上述精确放射治疗剂量计算及方案优化方法面临的新课题以及作者参与的中法合作项目“Optimization of Tumor Image Guided Radiotherapy, TIGRE”(项目编号: ANR-09-BLAN-0378-01)、国家自然科学基金国际合作与交流项目“精确放疗中的图像引导与生物优化问题研究”(项目编号: 60911130370), 针对笔形束剂量计算模型、生物模型在 IMRT 方案优化中的应用、IMRT 中随机几何不确定性的校正以及头颈部肿瘤自适应放射治疗中最佳重新规划的次数和时间等方面展开了较为深入的研究。

论文各章的内容具体安排如下:

第一章为绪论。

第二章在对经典笔形束算法进行研究的基础上, 提出一种新的笔形束剂量计算模型, 主要从剂量计算所使用的坐标系统、非均匀介质校正和快速傅立叶变换加速三方面进行了改进。新的笔形束算法需要模拟不同 DSS 的笔形束核。首先使用蒙特卡罗软件模拟笔形束核, 然后, 提取并存储不同 DSS 下笔形束核的深度信息和剖面分量的参数, 最后, 在不同的假设体模上验证新算法的有效性。

第三章主要研究了方案优化算法。在总结将生物准则转换成等效凸生物准则方法的基础上, 提出了利用等效凸生物准则构造目标函数的方法, 给出基于等效凸生物准则的子目标函数及其一阶导数, 并使用基于梯度的优化算法求解基于注量图的方案优化问题。新算法与基于剂量-体积准则的方案优化算法及其它两种基于 NTCP 准则的方案优化算法在 10 例前列腺肿瘤患者上进行了对比分析。

第四章通过在频域对已有的序列反卷积方法进行分析得知,反卷积核算法优于直接序列反卷积算法是因为反卷积算法等于直接序列展开算法和一个低通滤波器的乘积,可以进一步抑制反卷积结果中的高频信息。由于这个低通滤波器的性能与序列展开的截止项有关,其输出性能不能随便修改。因此,我们提出一种基于直接序列展开算法和频域滤波的混合反卷积方法。使用经典滤波器对直接序列反卷积结果进行滤波,并根据随机几何不确定性的运动幅度修改滤波参数,使得卷积后的注量图与静态注量图的差值尽可能的小。

第五章重点研究了离线自适应放射治疗方法,提出了一种确定离线自适应放射治疗中最佳重新规划时间及次数的方法。相对于直接根据体模内解剖结构或剂量变化确定最佳优化时间的方法,新方法根据累积剂量分布来确定最佳重新规划的时间及次数,综合地考虑了重新规划对整体剂量分布的影响,而不仅仅是对某次分次照射进行校正。根据是否考虑腮腺组织的过剂量照射,分别给出了两种情况下头颈部肿瘤最佳重新规划的时间及次数。

第六章对全文工作进行了总结,并对后续研究工作进行了展望。

1.4 参考文献

- [1] Lemoigne Y, Caner A. Radiotherapy and Brachytherapy [M]. France: Proceedings of the NATO Advanced Study Institute on Physics of Modern Radiotherapy & Brachytherapy Archamps, 2009:11.
- [2] 金雏凤, 图像引导放射治疗模式下精确放射治疗生物学效应研究[D],中国科学技术大学, 2003.
- [3] Bentzen SM, Heeren G, Cottier B, et al. Towards evidence-based guidelines for radiotherapy infrastructure and staffing needs in Europe: the ESTRO QUARTS project [J]. Radiother Oncol, 2005, 75:355-365.
- [4] 胡逸民主编.肿瘤放射物理学[M].北京:原子能出版社.1999:550.
- [5] 雒书朋, 王丽玲, 精确放射治疗的发展[J]. 医疗卫生设备, 2007, 28(11):57-58.
- [6] 翁学军, 放射治疗剂量计算的蒙特卡罗方法研究[D], 东南大学, 2003.
- [7] Boyer AL, Schultheiss T. Effects of dosimetric and clinical uncertainty on complication-free local tumor control [J]. Radiotherapy Oncology, 1988, 11:65-71.
- [8] ICRU, Determination of absorbed dose in a Patient irradiated by beams of X or Gamma rays in radiotherapy procedures[R], ICRU Report 24, 1976
- [9] Siebers JV, Tong S, Lauterbach M et al. Acceleration of dose calculations for intensity-modulated radiotherapy [J]. Med Phys, 2001, 28:903-910
- [10] Ahnesjö A, Aspradakis MM. Dose calculations for external photon beams in radiotherapy [J]. Phys Med Biol, 1999, 44:R99-R155.
- [11] Fogliata A, Vanetti E, Albers D, et al. On the dosimetric behaviour of photon dose calculation algorithms in the presence of simple geometric heterogeneities: comparison with Monte Carlo calculations [J]. Phys Med Biol, 2007, 52(5):1363-1385.
- [12] Mackie TR, Scrimger JW, Battista JJ. A convolution method of calculating dose for 15-MV x rays [J]. Med Phys, 1985, 12:188-96.
- [13] Mohan R, Chui C, Lidofsky L. Differential pencil beam dose computation model for photons [J]. Med Phys, 1986, 13:64-73.
- [14] Ahnesjö A, Andreo P, Brahme A. Calculation and application of point spread functions for treatment planning with high energy photon beams [J]. Acta Oncol, 1987, 26:49-56.
- [15] Ahnesjö A. Collapsed cone convolution of radiant energy for photon dose calculation in heterogeneous media [J]. Med Phys, 1989, 16:577-592.
- [16] Boyer AL, Mok EC. Calculation of photon dose distributions in an inhomogeneous medium using convolutions [J]. Med Phys, 1986, 13:503-509.

- [17] Erdi AK, Ellen DY, Murray HL, et al. Use of the fast Hartley transform for three-dimensional dose calculation in radionuclide therapy [J]. *Med Phys*, 1998, 25:2226-2233.
- [18] Lu W, Olivera GH, Chen ML, et al. Accurate convolution/superposition for multi-resolution dose calculation using cumulative tabulated kernels [J]. *Phys Med Biol*, 2005, 50(4):655-80.
- [19] Whitton K, Hu XS, Yu CX et al. An FPGA solution for radiation dose calculation [C]. *IEEE Symposium on Field Programmable Custom Computing Machines*, Napa, California, USA, 2006, pp. 227–236.
- [20] Zhou B, Hu XS, Chen DZ, et al. Hardware acceleration for 3-D radiation dose calculation [C]. *IEEE International Conference on Application-Specific Systems, Architectures and Processors*, Montreal, Quebec, Canada, 2007, pp. 290-295.
- [21] de Greef M, Crezee J, van Eijk JC, et al. Accelerated ray tracing for radiotherapy dose calculations on a GPU [J]. *Med Phys*, 2009, 36:4095-4102.
- [22] Hissoiny S, Ozell B, Després P. 2009 Fast convolution-superposition dose calculation on graphics hardware [J]. *Med Phys*, 36:1998-2005.
- [23] Hissoiny S, Ozell B, Després P. A convolution-superposition dose calculation engine for GPUs [J]. *Med Phys*, 2010, 37:1029-1037.
- [24] Ahnesjö A, Saxner M, Trepp A. A pencil beam model for photon dose calculation [J]. *Med Phys*, 1992, 19(2):263-73.
- [25] Liu HH, Mackie TR, McCullough EC. A dual source photon beam model used in convolution/superposition dose calculations for clinical megavoltage x-ray beams [J]. *Med Phys*, 1997, 24(12):1960-1974.
- [26] Yang Y, Xing L, Boyer AL, et al. A three-source model for the calculation of head scatter factors [J]. *Med Phys*, 2002, 29(9):2024-33.
- [27] Tillikainen L, Siljamäki S, Helminen H, et al. Determination of parameters for a multiple-source model of megavoltage photon beams using optimization methods [J]. *Phys Med Biol*, 2007, 52(5):1441-1467.
- [28] Chui CS, Mohan R. Extraction of pencil beam kernels by the deconvolution method [J]. *Med Phys*, 1988, 15:138–144.
- [29] Bortfeld T, Schlegel W, Rhein B. Decomposition of pencil beam kernels for fast dose calculation in three-dimensional treatment planning [J]. *Med Phys*, 1993, 20:311–318.
- [30] Ceberg CP, Bjarngard BE, Zhu TC. Experimental determination of the dose kernel in high-energy x-ray beams [J]. *Med Phys*, 1996, 23:505–511.
- [31] Storchi PRM, van Battum LJ, Woudstra E. Calculation of a pencil beam kernel from measured photon beam data [J]. *Phys Med Biol*, 1999, 44:2917–2928.
- [32] Bourland JD, Chaney EL. A finite-size pencil beam model for photon dose calculations in three dimensions [J]. *Med Phys*, 1992, 19:1401–1412.
- [33] Ostapiak Z, Zhu Y, van Dyk J. Refinements of the finite-size pencil beam model of three-dimensional photon dose calculation [J]. *Med Phys*, 1997, 24:743–750.
- [34] Jeleń U, Söhn M, Alber M. A finite size pencil beam for IMRT dose optimization [J]. *Phys Med Biol*, 2005, 50(8):1747-1766.
- [35] Jeleń U, Alber M. A finite size pencil beam algorithm for IMRT dose optimization: density corrections [J]. *Phys Med Biol*, 2007, 52:617–633.
- [36] Ulmer W, Pyyry J, Kaissl W. A 3D photon superposition/convolution algorithm and its foundation on results of Monte Carlo calculations [J]. *Phys Med Biol*, 2005, 50:1767–1790.
- [37] Tillikainen L, Helminen H, Torsti T, et al. A 3D pencil-beam-based superposition algorithm for photon dose calculation in heterogeneous media [J]. *Phys Med Biol*, 2008, 53(14):3821-3839.

- [38] Knöös T, Wieslander E, Cozzi L, et al. Comparison of dose calculation algorithms for treatment planning in external photon beam therapy for clinical situations [J]. *Phys Med Biol*, 2006, 51(22):5785-807.
- [39] Robinson D. Inhomogeneity correction and the analytic anisotropic algorithm [J]. *J Appl Clin Med Phys*, 2008, 9(2):112-122.
- [40] Gray A, Oliver LD, Johnston PN. The accuracy of the pencil beam convolution and anisotropic analytical algorithms in predicting the dose effects due to attenuation from immobilization devices and large air gaps [J]. *Med Phys*, 2009, 36(7):3181-3191.
- [41] Yoo S, Wu Q, O'Daniel J, et al. Comparison of 3D conformal breast radiation treatment plans using the anisotropic analytical algorithm and pencil beam convolution algorithm [J]. *Radiother Oncol*, 2012, 103(2):172-177.
- [42] Panettieri V, Barsoum P, Westermarck M, et al. AAA and PBC calculation accuracy in the surface build-up region in tangential beam treatments. Phantom and breast case study with the Monte Carlo code PENELOPE [J]. *Radiother Oncol*, 2009, 93(1):94-101.
- [43] Li X A, Yu C, Holmes T. A systematic evaluation of air cavity dose perturbation in megavoltage x-ray beams [J]. *Med Phys*, 2000, 27:1011-1017.
- [44] Nikos P, Jerry J, Battista AL, et al. Tissue inhomogeneity corrections for megavoltage photon beams. AAPM Report No. 85, Medical Physics Publishing, 2004.
- [45] Briemeister JF. MCNP-general Monte Carlo N-particle transport code version 4b. Report LA-12625-M, 1997.
- [46] Agostinelli S, Allison J, Amako K, et al., "Geant4 Collaboration," *Nucl. Instrum. and Methods in Phys. Res. A*, 2003, 506: 250-303.
- [47] Ferrari, A., P. Sala, A. Fassò. FLUKA: a multi-particle transport code(Program Verison 2011). 2011 (Milan: INFN)
- [48] Nelson WR, Hirayama H, Rogers DWO. The EGS4 code system. SLAC-Report-265, Stanford Linear Accelerator Center, 1985.
- [49] Rogers DWO, Faddegon BA, Ding GX, et al, BEAM: A Monte Carlo code to simulate radiotherapy treatment units [J]. *Med Phys*, 1995, 22(5):503-525,.
- [50] Salva F, Fernández-Varea JM, Sempau J. Penelope-A code system for Monte Carlo simulation of electron and photon transport [M]. Workshop Proceedings, Barcelona, Spain, 2006.
- [51] Hissoiny S, Ozell B, Bouchard H, et al. GPUMCD: a new GPU-oriented Monte Carlo dose calculation platform [J]. *Med Phys*, 2011, 38:754-764.
- [52] Jia X, Gu X, Graves Y J, et al. GPU-based fast Monte Carlo simulation for radiotherapy dose calculation [J]. *Phys Med Biol*, 2011, 56:7017-7031.
- [53] Jahnke L, Fleckenstein J, Wenz F, et al. GMC: A GPU implementation of a Monte Carlo dose calculation based on Geant4 [J]. *Phys. Med Biol*, 2012, 57:1217-1229.
- [54] Townson RW, Jia X, Tian Z, et al. GPU-based Monte Carlo radiotherapy dose calculation using phase-space sources [J]. *Phys Med Biol*, 2013, 58(12): 4341-4356.
- [55] Williams ML, Sajo E. Deterministic calculations of photon spectra for clinical accelerator targets [J]. *Med Phys*, 2002, 29(6): 1019-1028.
- [56] Williams ML, Ilas D, Sajo E, et al. Deterministic photon transport calculations in general geometry for external beam radiation therapy [J]. *Med Phys*, 2003, 30(12): 3183-3195.
- [57] Gifford KA, Horton JL, Wareing TA, et al. Comparison of a finite-element multigroup discrete-ordinates code with Monte Carlo for radiotherapy calculations [J]. *Phys Med Biol*, 2006, 51(9): 2253-2265.
- [58] Vassiliev ON, Wareing TA, McGhee J, et al. Validation of a new grid-based Boltzmann

- equation solver for dose calculation in radiotherapy with photon beams [J]. *Phys Med Biol*, 2010, 55(3): 581-598.
- [59] Fogliata A, Nicolini G, Clivio A, et al. Dosimetric validation of the Acuros XB Advanced Dose Calculation algorithm: fundamental characterization in water [J]. *Phys Med Biol*, 2011, 56(6): 1879-1904.
- [60] Bush K, Gagne IM, Zavgorodni S, et al. Dosimetric validation of Acuros XB with Monte Carlo methods for photon dose calculations [J]. *Med Phys*, 2011, 38(4): 2208-2221.
- [61] Allen Li X, Alber M, Deasy JO, et al. The use and QA of biologically related models for treatment planning: short report of the TG-166 of the therapy physics committee of the AAPM [J]. *Med Phys*, 2012, 39(3): 1386-1409.
- [62] Bortfeld T, Schmidt-Ullrich R, de Neve W, et al. *Image-guided IMRT* [M]. Springer, Berlin, 2005, pp. 199-216.
- [63] Boyer AL, Yu CX. Intensity-modulated radiation therapy with dynamic multileaf collimators[J]. *Semin Radiat Oncol*, 1999, 9(1): 48-59.
- [64] Dai J, Zhu Y. Minimizing the number of segments in a delivery sequence for intensity-modulated radiation therapy with multileaf collimator [J]. *Med Phys*, 2001, 28(10): 2113-2120.
- [65] Xia P, Verhey LJ. Multileaf collimator leaf sequencing algorithm for intensity modulated beams with multiple static segments [J]. *Med Phys*, 1998, 25(8):1424-1434.
- [66] Wu Q, Mohan R. Algorithms and functionality of an intensity modulated radiotherapy optimization system [J]. *Med Phys*, 2000, 27: 701-711.
- [67] Wu Q, Mohan R. Multiple local minima in IMRT optimization based on dose-volume criteria [J]. *Med Phys*, 2002, 29(7): 1514-1527.
- [68] Wu Q, Mohan R, Niemierko A, et al. Optimization of intensity-modulated radiotherapy plans based on the equivalent uniform dose [J]. *Int J Radiat Oncol Biol Phys*, 2002, 52: 224-235.
- [69] Choi B, Deasy JO. The generalized equivalent uniform dose function as a basis for intensity-modulated treatment planning [J]. *Phys Med Biol*, 2002, 47: 3579-3589.
- [70] Romeijn HE, Dempsey JF, Li JG. A unifying framework for multi-criteria fluence map optimization models [J]. *Phys Med Biol*, 2004, 49: 1991-2013.
- [71] Hoffmann AL, Den Hertog D, Siem AY, et al. Convex reformulation of biologically-based multi- criteria intensity-modulated radiation therapy optimization including fractionation effects [J]. *Phys Med Biol*, 2008, 53: 6345-6362.
- [72] Kalinowski T. Realization of intensity modulated radiation fields using multileaf collimators [J]. *Lecture Notes in Computer Science*, 2005, 21: 319-320.
- [73] Baatar D, Hamacher HW, Ehrigott M, et al. Decomposition of integer matrices and multileaf collimator sequencing [J]. *Discrete Appl Math*, 2005, 152: 6-34.
- [74] Engel K. A new algorithm for optimal multileaf collimator field segmentation [J]. *Discrete Appl Math*, 2005, 152(1): 35-51.
- [75] Cambazard H, O'Mahony E, O'Sullivan B. A shortest path-based approach to the multileaf collimator sequencing problem [J]. *Discrete Appl Math*, 2012, 160(1):81-99.
- [76] Men C, Romeijn HE, Taşkin ZC, et al. An exact approach to direct aperture optimization in IMRT treatment planning [J]. *Phys Med Biol*, 2007, 52(24):7333-7352.
- [77] Cassioli A, Unkelbach J. Aperture shape optimization for IMRT treatment planning [J]. *Phys Med Biol*, 2013, 58(2):301-318.
- [78] Shepard D M, Earl M A, Li X A, Naqvi S and Yu C Direct aperture optimization: a turnkey solution for step-and-shoot IMRT [J]. *Med Phys*, 2002, 29: 1007-18

- [79] Earl M, Afghan M, Yu C, Jiang Z, et al. Jaws-only IMRT using direct aperture optimization [J]. *Med Phys*, 2007, 34(1): 307-314.
- [80] Li Y, Yao J, Yao D. Genetic algorithm based deliverable segments optimization for static intensity-modulated radiotherapy [J]. *Phys Med Biol*, 2003, 48: 3353-3374.
- [81] Cotrutz C, Xing L. Segment-based dose optimization using a genetic algorithm [J]. *Phys Med Biol*, 2003, 48(18): 2987-2998.
- [82] Hardemark A, Liander H, Rehbindler H, et al. Direct machine parameter optimization with Ray Machine in Pinnacle. Raysearch Laboratories White Paper, 2003.
- [83] Romeijn HE, Ahuja RK, Dempsey JF, et al. A column generation approach to radiation therapy treatment planning using aperture modulation [J]. *SIAM J Optim*, 2005, 15: 838-862.
- [84] Carlsson F. Combining segment generation with direct step-and-shoot optimization in intensity-modulated radiation therapy [J]. *Med Phys*, 2008, 35(9): 3828-3838.
- [85] Preciado-Walters F, Langer MP, Rardin RL, et al. Column generation for IMRT cancer therapy optimization with implementable segments. *Ann Oper Res*, 2006 148: 65-79.
- [86] Men C, Jia X, Jiang SB. GPU-based ultra-fast direct aperture optimization for online adaptive radiation therapy [J]. *Phys Med Biol*, 2010, 55(15):4309-4319.
- [87] van Herk M. Errors and margins in radiotherapy [J]. *Semin Radiat Oncol*, 2004, 14: 52-64.
- [88] Yan D, Ziaja E, Jaffray D, et al. The use of adaptive radiation therapy to reduce setup error: a prospective clinical study [J]. *Int J Radiat Oncol Biol Phys*, 1998, 41(4): 715-720.
- [89] Hoogeman MS, van Herk M, de Bois J, et al. Strategies to reduce the systematic error due to tumor and rectum motion in radiotherapy of prostate cancer [J]. *Radiother Oncol*, 2005, 74(2): 177-185
- [90] Webb S. Limitations of a simple technique for movement compensation via movement-modified fluence profiles [J]. *Phys Med Biol*, 2005, 50(14): N155-161.
- [91] Keall PJ, Mageras GS, Balter JM, et al. The management of respiratory motion in radiation oncology report of AAPM Task Group 76 [J]. *Med Phys*, 2006, 33(10): 3874-3900.
- [92] Suh Y, Sawant A, Venkat R, et al. Four-dimensional IMRT treatment planning using a DMLC motiontracking algorithm [J]. *Phys Med Biol*, 2009, 54(12): 3821-3835.
- [93] Bortfeld T, Jokivarsi K, Goitein M, et al. Effects of intra-fraction motion on IMRT dose delivery: statistical analysis and simulation [J]. *Phys Med Biol*, 2002, 47(13): 2203-2220.
- [94] Lujan AE, Ten Haken RK, Larsen EW, et al. Quantization of setup uncertainties in 3-D dose calculations [J]. *Med Phys*, 1999, 26(11): 2397-2402.
- [95] McCarter SD, Beckham WA. Evaluation of the validity of a convolution method for incorporating tumor movement and setup variations into the radiotherapy treatment plans [J]. *Phys Med Biol*, 2000, 45(4): 923-931.
- [96] Craig T, Battista J, van Dyk J. Limitations of a convolution method for modeling geometric uncertainties in radiation therapy: I. The effect of shift invariance [J]. *Med Phys*, 2003, 30(8): 2001-2011.
- [97] Lu W, Olivera GH, Mackie TR. Motion-encoded dose calculation through fluence/sinogram modification [J]. *Med Phys*, 2005, 32(1): 118-127.
- [98] Lind BK, Kallman P, Sundelin B, et al. Optimal radiation beam profiles considering uncertainties in beam patient alignment [J]. *Acta Oncol*, 1993, 32: 331-342.
- [99] Löf J, Lind BK, Brahme A. Optimal radiation beam profiles considering the stochastic process of patient positioning in fractionated radiation therapy [J]. *Inverse Problems*, 1995, 11: 1189-1209.
- [100] Li JG, Xing L. Inverse planning incorporating organ motion [J]. *Med Phys*, 2000, 27(7): 1573-1578.

- [101] Unkelbach J, Oelfke U. Inclusion of organ movements in IMRT treatment planning via inverse planning based on probability distributions [J]. *Phys Med Biol*, 2004, 49: 4005–4029.
- [102] Unkelbach J and Oelfke U. Incorporating organ movements in IMRT treatment planning for prostate cancer: minimizing uncertainties in the inverse planning process [J]. *Med Phys*, 2005, 32(8): 2471–2483.
- [103] Chu M, Zinchenko Y, Henderson SG, et al. Robust optimization for intensity modulated radiation therapy treatment planning under uncertainty [J]. *Phys Med Biol*, 2005, 50(23): 5463–5477.
- [104] Baum C, Alber M, Birkner M, et al. Robust treatment planning for intensity modulated radiotherapy of prostate cancer based on coverage probabilities [J]. *Radiother Oncol*, 2006, 78(1): 27–35.
- [105] Olafsson A, Wright SJ. Efficient schemes for robust IMRT treatment planning [J]. *Phys Med Biol*, 2006, 51(21): 5621–5642.
- [106] Chan TC, Bortfeld T, Tsitsiklis JN. A robust approach to IMRT optimization [J]. *Phys Med Biol*, 2006, 51(10): 2567–2583.
- [107] Bortfeld T, Chan TCY, Trofimov A, et al. Robust management of motion uncertainty in intensitymodulated radiation therapy [J]. *Oper Res*, 2008, 56: 1461–1473.
- [108] Fan Y, Nath R. Intensity modulation under geometrical uncertainty: a deconvolution approach to robust fluence [J]. *Phys Med Biol*, 2010, 55(14): 4029–4045.
- [109] Castadot P, Lee JA, Geets X, et al. Adaptive radiotherapy of head and neck cancer [J]. *Semin Radiat Oncol*, 2010, 20(2): 84–93.
- [110] Yan D, Vicini F, Wong J, et al. Adaptive radiation therapy [J]. *Phys Med Biol*, 1997, 42(1): 123–132.
- [111] Schlegel WC, Bortfeld T, Grosu, AL, et al. New technologies in radiation oncology [M]. *Medical Radiology*, Berlin: Springer, 2006, pp. 321–336.
- [112] Wu C, Jeraj R, Olivera GH, et al. Re-optimization in adaptive radiotherapy [J]. *Phys Med Biol*, 2002, 47: 3181–3195.
- [113] Rehbrinder H, Forsgren C, Löf J. Adaptive radiation therapy for compensation of errors in patient setup and treatment delivery [J]. *Med Phys*, 2004, 31(12): 3363–3371.
- [114] de la Zerda A, Armbruster B, Xing L. Formulating adaptive radiation therapy (ART) treatment planning into a closed-loop control framework [J]. *Phys Med Biol*, 2007, 52(14): 4137–4153.
- [115] Sir MY, Epelman MA, Pollock SM. Stochastic programming for off-line adaptive radiotherapy [J]. *Ann Oper Res*, 2012, 196: 767–797.
- [116] Yan D, Liang J. Expected treatment dose construction and adaptive inverse planning optimization: implementation for offline head and neck cancer adaptive radiotherapy [J]. *Med Phys*, 2013, 40(2): 021719–10.
- [117] Mohan R, Zhang X, Wang H, et al. Use of deformed intensity distributions for on-line modification of image-guided IMRT to account for interfractional anatomic changes [J]. *Int J Radiat Oncol Biol Phys*, 2005, 61(4): 1258–1266.
- [118] Court LE, Dong L, Lee AK, et al. An automatic CT-guided adaptive radiation therapy technique by online modification of multileaf collimator leaf positions for prostate cancer [J]. *Int J Radiat Oncol Biol Phys*, 2005, 62(1): 154–163.
- [119] Feng Y, Castro-Pareja C, Shekhar R, et al. Direct aperture deformation: an interfraction image guidance strategy [J]. *Med Phys*, 2006, 33(12): 4490–4498.
- [120] Ahunbay E, Peng C, Chen G, et al. An on-line replanning scheme for interfractional

variations [J]. Med Phys, 2008, 35(8): 3607-3615

[121] Lei Y, Wu Q. A hybrid strategy of offline adaptive planning and online image guidance for prostate cancer radiotherapy [J]. Phys Med Biol, 2010, 55(8): 2221–2234.

[122] Li T, Thongphiew D, Zhu X, et al. Adaptive prostate IGRT combining online re-optimization and re-positioning: a feasibility study [J]. Phys Med Biol, 2011, 56(5): 1243-1258.

第二章 一种新的笔形束光子剂量计算方法

2.1 引言

在 IMRT 放射治疗中, 通常将大射野(broad field)分割成小的射野单元, 这些小射野单元被称为射束(beamlet)。每一个射束都对应一个复杂的注量, 即在该射束方向通过的所有粒子。大射野的剂量分布可以由这些射束的剂量分布卷积、叠加得到。这些射束的剂量分布需要在基于注量图的 IMRT 优化开始之前计算并存储。在优化过程中, 通过调整这些射束的注量分布使得靶区的剂量分布满足处方剂量要求而周围正常组织接收到的剂量尽可能的少。在优化结束后, 根据这些射束的注量分布, 使用更加精确的剂量算法计算最终的剂量分布, 如蒙特卡罗(Monte Carlo, MC)方法。这种基于预先计算射束剂量分布的 IMRT 优化方法需要满足以下三个要求: 1) 速度, 需要计算大量射束的剂量分布, 2) 简单的调试过程, 应该是在基于大射野测量的理想情况下进行调试, 3) 两种算法精度的接近程度, 在优化过程中使用精度低、速度快的剂量计算方法而最终剂量使用高精度的算法计算。

精确快速的 3D 剂量计算是放射治疗技术中的关键部分, 特别是在 IMRT 中。在实施方案优化和评价方案质量的时候, 物理师和医生非常依赖剂量计算的精度。逆向方案优化还需要剂量计算的速度足够快, 因为逆向方案优化过程是一个迭代的过程, 在这个过程中剂量计算要重复 10 到 1000 次^[1]。基于点核的卷积叠加算法和蒙特卡罗方法能够比较精确地计算剂量分布, 但是其所需计算时间比较长, 一般用来计算最终的剂量分布。而笔形束算法在非均匀介质中剂量计算的精度比较低, 但是由于其计算速度快的特点, 通常在 IMRT 中被用来预先计算剂量分布。使用笔形束核进行剂量计算可以理解为对不同光子注量强度的笔形束核进行卷积或叠加。已有大量学者对笔形束算法的精度和速度提出了改进^[2-8]。为了加快笔形束算法的计算速度, 一些学者使用快速变换的方法加速剂量计算, 如快速傅立叶变换(fast Fourier transform convolution, FFTC)^[7]和快速哈特莱变换(fast Hartley transform)^[8]。

使用基于笔形束核模型的剂量计算方法, 需要考虑以下三方面的因素^[9]: 1) 射线源能谱和几何结构; 2) 非均匀介质(人体)中的剂量计算; 3) 剂量计算所需时间。在本章中, 主要围绕剂量计算所使用坐标系统、非均匀介质中剂量计算和 FFTC 加速三个方面展开了较为深入的研究。

来自射线源的射线是发散的。早期的笔形束算法假设来自射线源的射线是相互平行的, 并按平行笔形束核直接计算 3D 剂量分布。这种方法忽略了由于射线发散所带来的误差, 在较小的源皮距下, 该误差将会超过 3%^[9]。然而直接在直角坐标系统中校正这类误差, 将笔形束核按射线发散角度旋转, 却增加了计算量, 延长了剂量计算所需时间。一些学者提出在球坐标系统下计算剂量分布, 如扩散坐标系统^[4, 6]和 beam's eye view 坐标系统^[10, 11]。由于球坐标系统的特殊性质, 这种方法可以避免旋转笔形束核并有效地校正由射线发散带来的误差, 而且没有增加过多的计算时间。但是在传统等角度采样方式下的球坐标系统中, 体素的体积随着到圆心的距离的增加而增大, 因而影响了剂量计算的精度。

笔形束核描述的是一束射线入射在半无限大均匀介质中的能量扩散模式, 其计算过程可以分成两步来完成: 1) 深度方向分量计算, 即该射束在深度方向上的每个平面上沉积的总能量; 2) 侧向分量计算, 即该射束在每个平面上沉积的总能量的扩散分布。因此, 在经典的笔形束模型中, 剂量计算所采用的校正算法也主要是沿深度方向和侧向方向分别进行校正的。在校正过程中, 由于所采用笔形束模型自身的结构特点也会引入一些误差。如在 AAA 算法中, 深度方向分量的校正以射束在体模表面的入射点为参考点进行的。而在侧向分量校正的时候, 侧向分量需要乘以一个包含深度信息的校正因子^[4, 6, 12]。因此, AAA 算法在进行侧向分量校正时, 需要考虑前面几层的密度信息。这种校正方法是一种近似计算, 降低了剂量计算的精度。

为了减少运算所需时间, FFTC 技术被用来加快剂量计算速度。但是由于笔形束模型本身

的结构特点及所采用的校正方法,限制了 FFTC 技术的使用。如在 AAA 算法中,由于深度方向分量的校正以射束在体模表面的入射点为参考点,在深度方向分量进行校正后,在同一球壳上的剂量计算点将不在同一球壳上;侧向分量中包含有深度信息,而且很难将这些深度信息分离出来。因此,在考虑深度方向分量校正的情况下,是不可以使用 FFTC 技术对 AAA 算法进行加速。

本章主要针对校正射线发散所使用的坐标系统和 AAA 算法在非均匀介质中的剂量校正以及用 FFTC 技术加速中存在的问题,提出了一种新的笔形束剂量计算模型。在新的笔形束模型中,采用等弧长的采样方式,减小了由于不同体积的体素所引入的误差。在非均匀介质校正中的剂量计算,以射束碰撞点为参考点,移动射束入射点的方法进行校正。这种校正方法可以将侧向分量中的深度信息分离出来,减少了由校正因子所带来的误差。同时,采用这种校正方法,同一球壳上的剂量计算点位置不变,因此可以直接使用 FFTC 技术进行加速。

2.2 笔形束剂量计算方法

笔形束是指零射野条件下由放射源发出的一束无穷小的射线束。然而在实际条件下无法达到零射野照射的要求,因此实际中使用的笔形束一般都是有尺寸大小的“窄束”。笔形束剂量计算方法认为:任何射野的电离辐射(ionizing radiation)在均匀介质中的剂量分布,都可以由很多相似的“窄束”产生的剂量分布叠加而成^[13]。这种通过对笔形束叠加得到总射野剂量分布的概念被广泛应用在放射治疗剂量计算中。笔形束卷积核是描述一束纤细射线入射到无限大均匀介质中的能量扩散模式。笔形束核可用 MC 方法模拟得到或对实验测量数据进行反卷积得到。为了提高笔形束剂量计算的精度,笔形束核一般分为原射线剂量和散射线剂量两部分,一般表示为指数函数^[2, 6]或高斯函数^[4]以及这两种函数的推广模型(几个指数函数或高斯函数和的形式)。利用笔形束核进行剂量计算的公式为:

$$D(r) = \int_E \iint_S \psi_E(s) \cdot h_E(z, r-s) ds dE \quad (2.1)$$

式中 $D(r)$ 为体模内 r 点处的剂量分布, ψ 为在该射野方向体模表层的注量分布, E 为入射射束的能谱, h_E 为预先计算好的笔形束核。由于非均匀介质会影响入射射线和次级电子的传输过程,因此在叠加过程中需要对笔形束核进行相应的校正。与三维卷积模型相比,笔形束模型预先做了一维的积分而简化为二维积分,从而大大地削减了计算量,因此在商业放射治疗系统中被广泛使用。

基于核模型的剂量计算方法是在一些模型假设的基础上进行剂量计算的,如射线源几何结构及能谱变化、非均匀介质中剂量计算、剂量卷积/叠加的方法(主要影响到计算时间)等。

1) 射线源模型

剂量计算的精度很大程度上与精确地模拟真实的治疗头有关。不同能量的笔形束核是互不相同的,因此总的剂量分布是不同能级下的剂量分布之和,如公式 2.1 所示。由于射线源发出射线的能谱是连续的,需要离散化之后才能在计算机上进行计算。将射线源假设为单能射线又会导致较大的剂量计算误差。Boyer 等人发现使用五个能级的单能射线足够将西门子 6MeV 治疗头发射线的能谱进行离散化^[14]。而且射线在输运过程中其能谱是变化的:1) 射束硬化,射束通过被照射物体时低能级射线首先被滤除;2) 离轴变软,随着离轴距离的增加由于射线通过平整滤波器被硬化,而导致射线能谱在侧向发生变化。不仅需要对原射线进行模拟,还需要考虑部分原射线与平整器、准直器碰撞产生的散射线以及在空气中的电离污染等。绪论中已经提到了一些学者将真实治疗头模拟成多源模型的方法。

2) 射束发散

对于发散射束,随着入射深度的增加,原射线注量按平方反比因子减少,而射野面积按线性增大。Papanikolaou 等人首先利用平行射束计算剂量分布,然后使用入射深度的平方反比因子校正体模内每点处的剂量分布^[15]。实验表明这种方法可以近似地校正由射线发散引入的误差。Sharpe 和 Battista 研究了直接使用平行射束进行剂量计算所引起的误差^[16]。实验表

明,在大射野、小的源皮距和高能射线的情况下,忽略射束发散会引入大概 3% 的误差。在直角坐标系统中,可以直接将笔形束核按射线发散角度旋转来校正这类误差^[17],但是由于这种方法增加了计算量,延长了剂量计算所需时间。对此,一些学者利用球坐标系统的特殊性质,在球坐标系统下计算剂量分布^[4, 6, 10, 11]。这种方法可以避免旋转笔形束核并有效地校正由射线发散带来的误差,而没有增加过多的计算时间。

3) 非均匀介质中的剂量校正

光子射束在体模内发生碰撞产生的初级散射光子在能量沉积点沉积的能量,与射束碰撞点和能量沉积点之间的物质有关。而由于大部分的多级散射光子是沿射束方向前行的,所以其能量沉积不仅与射束碰撞点和能量沉积点之间的物质有关,还与散射线初始方向有关。在基于核模型的剂量计算方法中,等效路径(equivalent path length, EPL)的方法通常被用来校正非均匀介质中的剂量分布^[18]。Seco 等人证明了基于相对电子密度的等效路径方法精度要高于基于密度的等效路径方法^[19]。在非均匀介质中,基于电子密度的射线等效路径方法可以定义为对射线通过的每一个体素进行拉伸:

$$d_{\text{eff}}(\mathbf{X}) = \int_{\mathbf{X}} \rho^{\text{elec}}(\mathbf{p}) / \rho_{\text{water}}^{\text{elec}} d\mathbf{p} \quad (2.2)$$

其中 \mathbf{X} 是射线实际通过路径, $\rho^{\text{elec}}(\mathbf{p})$ 是在点 \mathbf{p} 处的局部电子密度, $\rho_{\text{water}}^{\text{elec}}$ 是水的电子密度。

4) 卷积/叠加方法

尽管计算机技术持续快速发展和运算速度日益增加,但是直接叠加每个射束的笔形束核还是非常耗时的,阻碍了剂量计算方法在临床上的应用。一些学者使用特殊的硬件对这种直接叠加的方法进行了加速^[20]。尽管光子的输运过程是相互独立的,由于所考虑的计算系统日益复杂且无统一的标准,仍然阻碍了这种利用平行结构进行运算的方法在商用放射治疗系统中的应用。自从在方案优化算法的迭代过程中开始使用剂量计算方法后,出现了一些加快剂量计算的方法,主要有:直接叠加算法、快速算法^[7, 8]、筒串算法^[21]和混合算法^[22]。由于直接叠加算法的计算量大,所需计算时间长,一些学者使用快速傅立叶变换^[7]或快速哈特莱变换^[8]对直接叠加的方法进行加速。快速变化法也可以称为卷积的方法,因为在叠加的过程中没有考虑非均匀介质的影响,笔形束核是保持不变的。使用快速算法,忽略了非均匀介质对剂量分布的影响,严重地降低了剂量计算的精度。对此,Ahnesjö 等人提出筒串卷积算法^[21],在减少计算量的同时,还能有效地校正非均匀介质对剂量分布的影响。由于考虑了非均匀介质中的剂量校正,筒串卷积算法需要根据局部电子密度的分布对笔形束核进行拉伸。因此,筒串卷积算法只能采用叠加算法计算总的剂量分布,不能使用快速算法加速。为了进一步加快筒串卷积算法的计算速度,一些学者提出了混合算法,即在密度变化较大的区域使用筒串卷积算法,在密度变化平缓的区域使用快速算法加速剂量计算^[22]。

2.3 新的笔形束剂量计算方法

2.3.1 获取笔形束核参数

2.3.1.1 改进的坐标系统

在本章中,所有的射线被认为是来自于一个点源,这个点源被置于坐标系统的原点 O 处。所有的射线由原点向四周发散。相对于射束的中心轴,这些发散射线的偏转角度是各不相同的。在传统笔形束算法中,只模拟得到射野中心轴处的笔形束核。在校正射线发散的剂量计算过程中,其它发散射线的笔形束核是根据发散射线的偏转角度,适当地旋转中心轴处的笔形束核得到^[13]。但是在直角坐标系统中,为每条发散射线旋转笔形束核的方法是非常耗时的。因此,我们在球坐标系统下计算剂量分布,以减少由于旋转笔形束核所增加的计算量。如图 2-1 所示,射线源在球坐标系统的原点处,来自射线源的射线垂直于每一个球壳。这里的球壳表示球坐标系统中深度方向的每一层。射线在某球壳处发生碰撞释放的能量被认为在该球壳

上以碰撞点为中心向四周传播，如图 2-1 所示。因此，可以按平行射束的方式在球坐标系中计算剂量分布，从而避免了在直角坐标系统中对笔形束核的旋转。

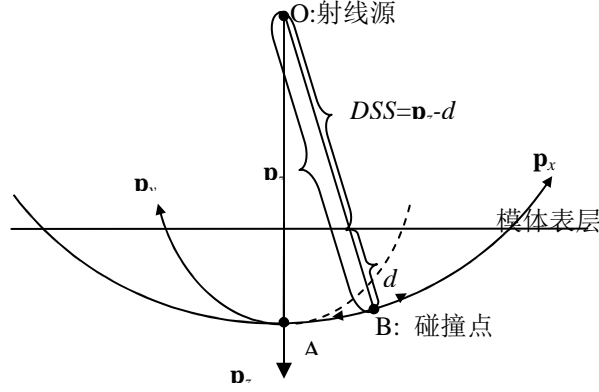


图 2-1 球坐标系：点 A 和点 B 是深度为 p_z 的球壳上的两个点，DSS 表示通过点 B 的射线的射线源到皮肤表面的距离， d 是该射线在体模内的传输路径。

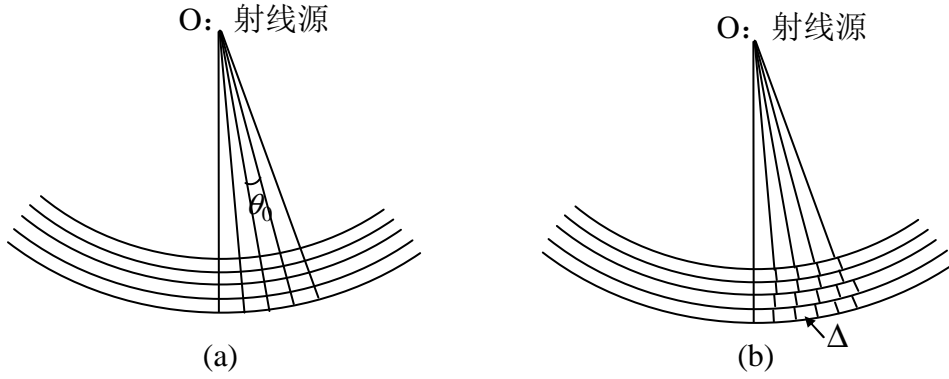


图 2-2 球坐标系中的两种采样方式：(a) 按固定角度采样；(b) 按固定弧长采样。

笔形束核的坐标系是以一个正交基的形式给出。原点在射线源处，坐标系统 z 轴的正半轴通过体模中的等中心点(isocenter)。X 轴和 y 轴分别与初级准直器的两条轴线对齐。通常笔形束核是按固定顶角进行采样的，如 AAA 算法。在这种采样方式下，由直角坐标系统到球坐标系统的映射关系可以表示为：

$$\mathbf{x} \mapsto \left(\frac{\arctan(\mathbf{x}_x / \mathbf{x}_z)}{\theta_0}, \frac{\arctan(\mathbf{x}_y / \mathbf{x}_z)}{\theta_0}, \sqrt{\mathbf{x}_x^2 + \mathbf{x}_y^2 + \mathbf{x}_z^2} \right) := \mathbf{p} \quad (2.3)$$

其中 $\mathbf{x} = (\mathbf{x}_x, \mathbf{x}_y, \mathbf{x}_z)$ 表示直角坐标系统中的一点，在球坐标系统中对映该点的位置表示为 $\mathbf{p} = (\mathbf{p}_x, \mathbf{p}_y, \mathbf{p}_z)$ ， θ_0 表示顶角的采样角度(如图 2-2(a)所示)。但是，按固定的顶角 θ_0 进行采样，体素的体积会随着离圆心的距离的增加而逐渐变大，降低了剂量计算精度。为了避免由不同体积的体素所引起的误差，在本章中我们按固定弧长进行采样(如图 2-2(b)所示)，由直角坐标系统到球坐标系统的新映射关系 $M: \mathcal{R}_+^3 \rightarrow \mathcal{R}_+^3$ 可以表示为：

$$\mathbf{x} \mapsto \left(\frac{\arctan(\mathbf{x}_x / \mathbf{x}_z)}{\Delta} \mathbf{p}_z, \frac{\arctan(\mathbf{x}_y / \mathbf{x}_z)}{\theta_0} \mathbf{p}_z, \mathbf{p}_z \right) := \mathbf{p} \quad (2.4)$$

其中 Δ 的值是固定的，表示按固定的弧长进行采样， $\mathbf{p}_z = \sqrt{\mathbf{x}_x^2 + \mathbf{x}_y^2 + \mathbf{x}_z^2}$ 。

2.3.1.2 笔形束核获取

单能笔形束核函数 $h_E(\mathbf{x})$ 被认为是用来描述一束能量为 E 的单能射线垂直入射到半无限大水体模内产生的能量分布。单能笔形束核可以使用 MC 方法精确模拟得到^[23]。每一个单能笔形束核 h_E 需要模拟一千万个单能粒子的剂量分布得到，以保证峰值剂量区域中的标准差小于 0.3%。在本章中，非均匀介质中的剂量校正是以射束在体模中的碰撞点为参考进行的，因此需要模拟不同射线源到皮肤表层距离(distance from source to surface, DSS)下的笔形束核。为了能在改进后的球坐标系统中进行剂量计算，在直角坐标系统下获取的笔形束核必须转换到改进后的球坐标系统中。

为了减少运算时间，我们在剂量计算的过程中直接使用多能笔形束核进行笔形束核卷积/叠加。多能笔形束核可以表示为对不同单能笔形束核加按能谱分布进行加权的和：

$$h(\mathbf{x}) = \frac{\int h_E(\mathbf{x}) S(E) dE}{\int S(E) dE} \quad (2.5)$$

其中 $S(E)$ 表示所使用多能射线的能谱。该笔形束核可以根据公式 2.3 中的映射关系直接投影到改进后的球坐标系统中。在改进后的球坐标系统中，不同 DSS 的多能笔形束核可以由下式得到：

$$h_{DSS}(\mathbf{p}) = h(\mathbf{x}) \det(J(M(\mathbf{x}))) \quad (2.6)$$

其中 $\mathbf{x} = M^{-1}(\mathbf{p})$ 表示在改进后球坐标系统中的点 \mathbf{p} 在直角坐标系统中的对应位置， $J(M(\mathbf{x}))$ 是两坐标系统之间映射关系 M 的雅克比行列式，雅克比行列式的值 $\det(J(M(\mathbf{x})))$ 用来表示两不同坐标系统中同一位置处体素体积之间的关系：

$$\det(J(M)) = \frac{\mathbf{x}_z(\mathbf{x}_x^2 + \mathbf{x}_y^2 + \mathbf{x}_z^2)^{3/2}}{(\mathbf{x}_x^2 + \mathbf{x}_z^2)(\mathbf{x}_y^2 + \mathbf{x}_z^2)} \quad (2.7)$$

2.3.1.3 笔形束核参数提取

笔形束函数通常表示为距离体模表层深度和到射束中心轴半径长度的能量沉积函数。为了缩短剂量计算的运算时间，笔形束函数中的参数存储在查找表中以便运算过程中调用。通常，将笔形束核分为两部分：深度方向分量和侧向分量。在本章中，深度方向分量表示该射束在体模中深度为 \mathbf{p}_z 的球壳上释放的所有能量：

$$I_{DSS}(\mathbf{p}_z) = \Phi \iint h_{DSS}(\mathbf{p}_x, \mathbf{p}_y, \mathbf{p}_z) d\mathbf{p}_x d\mathbf{p}_y \quad (2.8)$$

其中 Φ 表示在该射野方向上体模表层的能量注量分布， h_{DSS} 表示不同 DSS 的笔形束核。侧向分量用来描述射束在每一个球壳上释放的能量在对应球壳上的散射传播。在剂量计算过程中，我们使用筒串卷积方法进行剂量计算。因此，在每一层球壳上按极坐标进行重新采样得到 $h_{DSS}(r, \mathbf{p}_z)$ 。侧向分量 $f_{DSS}(r, \mathbf{p}_z)$ 表示在距离射束中心轴长度为 r 的体素中接收到的部分由该射束在深度为 \mathbf{p}_z 的球壳上释放的能量。侧向分量可以从笔形束核中提取得到：

$$f_{DSS}(r, \mathbf{p}_z) = h_{DSS}(r, \mathbf{p}_z) / I_{DSS}(\mathbf{p}_z) \quad (2.9)$$

由于直接存储侧向分量需要大量的存储空间，侧向分量一般表示为关于离轴半径 r 的函数。使用拟合算法可以从式 2.9 中得到侧向分量函数中各参数的值。在本章中我们使用 FFTC 技术对剂量计算进行加速，而使用 FFTC 技术需要笔形束函数中的深度信息和侧向信息相互独立。因此，在本章中我们使用了由 Tillikainen 等人^[6]提出的侧向散射函数来表示侧向分量。该侧向散射函数表示为包含深度信息的权重参数 $c_{DSS,i}(\mathbf{p}_z)$ 和包含侧向散射信息的指数函数的乘积和的形式，易于将侧向信息和深度信息分离开：

$$k_{DSS}(r, p_z) = \sum_{i=1}^{N=6} c_{DSS,i}(p_z) \frac{1}{\mu_i} e^{-\mu_i r} \quad (2.10)$$

其中衰减参数 μ_i 在所有深度的球壳中都是相同的，以减少参数个数，方便在剂量计算过程中使用。使用最小二乘法进行拟合，可以从公式 2.9 中得到公式 2.10 中参数 $c_{DSS,i}(\mathbf{p}_z)$ 的值。参数 μ_i 的取值是在 $1/\mu_i$ 的有效取值范围(1mm 到 200mm)内按相等的对数间隔进行取值。最后，将不同 DSS 的笔形束核中的深度分量 $I_{DSS}(\mathbf{p}_z)$ ，衰减因子 μ_i 和侧向散射函数中的权重参数 $c_{DSS,i}(\mathbf{p}_z)$ 存入查找表，以便在剂量计算过程中使用。

2.3.2 叠加剂量计算方法

为了增加非均匀介质中剂量计算的精度，本章中使用叠加的方法来计算体模内每点处所有沉积的能量。使用叠加类的方法易于来使用等效路径^[24]的方法来校正非均匀介质中的剂量分布。在本章中，我们使用 Ahnesjö^[21]提出的筒串卷积算法进行笔形束核叠加。筒串卷积算法将体模内的空间以射束与球壳的交点为中心划分为具有一定立体角的同轴锥形筒串，由交点处释放的所有能量沿筒串中心轴性进行传输，并在筒串中心轴上的体积单元内衰减和沉积。下面将详细地介绍在不同介质中的笔形束叠加算法。

2.3.2.1 均匀介质中的剂量计算

众所周知，粒子在体模内的传输路径是不确定的。对于笔形束剂量算法，粒子在体模内的输运过程被认为是分两步来实现的：首先粒子被认为沿着射束的中心轴到达深度为 \mathbf{p}_z 的球壳层，部分粒子在射束中心轴与球壳的交点发生碰撞并释放能量；然后释放的能量以交点为中心，沿有限的几个筒串中心轴向周围散射^[6]。在均匀介质中，由一束笔形束射束在体模内 \mathbf{p} 点处沉积的能量 $E(\mathbf{p})$ 可以表示为深度分量 $I_{DSS}(\mathbf{p}_z)$ 和侧向分量 $k_{DSS}(r, \mathbf{p}_z)$ 乘积的形式：

$$E(\mathbf{p}) = I_{DSS}(\mathbf{p}_z) k_{DSS}(r, \mathbf{p}_z) \quad (2.11)$$

其中深度分量 $I_{DSS}(\mathbf{p}_z)$ 为该射束在交点处释放的所有能量(式 2.8)，侧向分量 $k_{DSS}(r, \mathbf{p}_z)$ 为部分在交点释放的能量在点 \mathbf{p} 处沉积(式 2.10)， r 是从交点到点 \mathbf{p} 的距离。由于射束的入射角度和体模表面的凹凸不平，公式 2.11 中的 DSS 为自变量，可以通过 \mathbf{p}_z-d 计算得到(如图 2-1 所示)，其中 d 为等效深度，其值可以通过射线跟踪算法计算沿射束中心轴从与体模表层的交点到与球壳交点之间所有体素的等效长度得到。公式 2.11 中其余的参数都可以在前文提到的查找表中找到。

2.3.2.2 非均匀介质中的剂量计算

由于笔形束核一般被分为深度分量和侧向分量两部分，非均匀介质中的剂量校正一般是通过分别对这两部分分量进行校正来实现的。在经典的笔形束算法中，深度分量的校正通过移动射束中心轴与球壳的交点和剂量计算点来实现的(如图 2-3(a)所示)。但是在球坐标系中，这种校正方法会引入一些误差。如图 2-3(a)所示，这种方法在移动交点 \mathbf{B} 的时候，点 \mathbf{B} 所在球壳上的所有点也随着 \mathbf{B} 点的移动而改变位置。由于 \mathbf{B} 点所在球壳和 \mathbf{B}' 所在球壳的曲率不同且弧长 \mathbf{AB} 大于弧长 $\mathbf{A'B'}$ ，当点 \mathbf{B} 移到到点 \mathbf{B}' 时，点 \mathbf{A} 没有移动到 $\mathbf{A'}$ 点，而是移动到 $\mathbf{A''}$ 点。而在笔形束核的模拟过程中，只提取了弧 \mathbf{AB} 所在球壳和弧 $\mathbf{A'B'}$ 所在球壳的笔形束核参数，而没有提取弧 $\mathbf{A''B'}$ 所在球壳的笔形束核参数。因此在剂量计算过程中直接使用弧 $\mathbf{A'B'}$ 所在球壳的笔形束核参数代替弧 $\mathbf{A''B'}$ 所在球壳的笔形束核参数会给剂量计算引入一些误差。在 AAA 算法中使用了这种近似代替的方法并对弧长 \mathbf{AB} 的等效路径乘以校正因子^[6]。在本章中，为了避免这种误差的出现，通过移动射束中心轴和体模表层的交点位置和改变射束的 DSS 来实现对深度分量的校正。新的校正方法在校正深度分量的时候，射束中心轴和球壳交点 \mathbf{B} 的位置没有改变，因此点 \mathbf{B} 所在球壳的位置不用随之改变，从而避免了由于球壳位置改变而引入的误差。

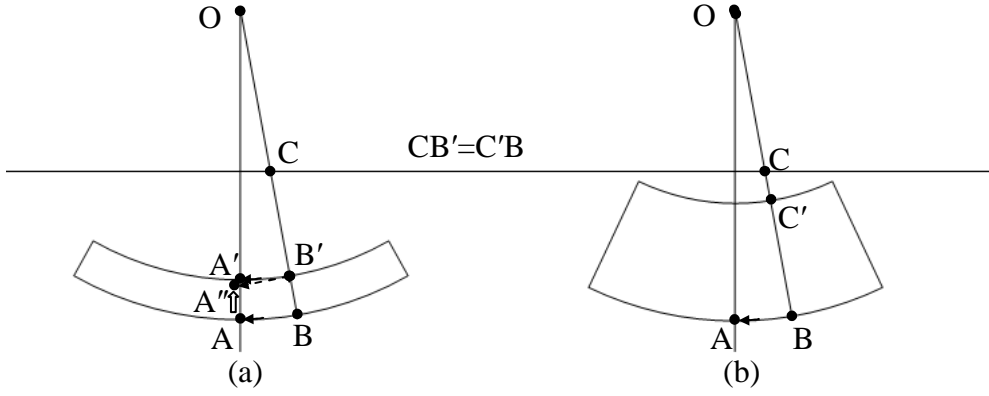


图 2-3 非均匀介质中两种不同的剂量校正方法：(a) 改变射束中心轴与球壳交点 B 的位置；(b) 改变射束中心轴与体模表层交点的位置，同时改变该射束的 DSS。

因此，在新的笔形束算法中，深度方向分量 $I_{DSS}(\mathbf{p}_z)$ 是通过改变射束中心轴和体模表层交点的位置和改变该射束的 DSS 来实现的。改变后的 DSS 值可以根据公式 $DSS = \mathbf{p}_z - deff(P)$ 计算得到，其中 P 为沿射束中心轴从体模表层到与球壳交点 \mathbf{p}' 之间的实际路径， $deff(P)$ 为 P 的等效路径。深度方向分量 $I'_{DSS}(\mathbf{p}_z)$ 还与射束中心轴和该球壳交点 \mathbf{p}' 处的局部电子密度有关系，其公式为：

$$I'_{DSS}(\mathbf{p}_z) = I_{DSS}(\mathbf{p}_z) \rho_w(\mathbf{p}') \quad (2.12)$$

其中 $\rho_w(\mathbf{p}')$ 是点 \mathbf{p}' 处的相对局部电子密度。

侧向分量的校正同样也是使用等效路径的方法进行的。与深度方向分量的校正不同，侧向分量的等效路径是沿筒串中心轴对从射束中心轴与球壳的交点 \mathbf{p}' 到实际计算剂量的点 \mathbf{p} 之间的实际路径 $C(\mathbf{p})$ 进行拉伸得到的。在交点处释放的能量 $I'_{DSS}(\mathbf{p}_z)$ 沿有限的筒串中心轴进行散射传播。而沿实际路径 $C(\mathbf{p})$ 所在筒串中心轴传递的能量，在经过实际路径 $C(\mathbf{p})$ 上介质的衰减后，最终在剂量计算点 \mathbf{p} 处沉积的能量 $k'_{DSS}(r, \mathbf{p}_z)$ 为：

$$k'_{DSS}(r, \mathbf{p}_z) = k_{DSS}(r', \mathbf{p}_z) \rho_w(\mathbf{p}) \quad (2.13)$$

其中 $r' = deff(C(\mathbf{p}))$ 为等效半径长。与 AAA 算法相比，由于新算法在剂量校正过程中没有改变射束中心轴与球壳交点的位置，侧向分量的校正不需要乘以包含深度信息的校正因子。

分别对笔形束算法中的深度方向分量(公式 2.12)和侧向分量(公式 2.13)进行校正后，由一束射束在剂量计算点 \mathbf{p} 处沉积的能量为：

$$E(\mathbf{p}) = I'_{DSS}(\mathbf{p}_z) k'_{DSS}(r, \mathbf{p}_z) \quad (2.14)$$

在 \mathbf{p} 点处沉积的所有能量为射野内所有射束在该点沉积的能量之和：

$$E_{tot}(\mathbf{p}) = \iint_S E(\mathbf{p}) dS \quad (2.15)$$

其中 S 为射野内的所有射束。为了能在计算机上实现算法，能量叠加需要在有限的离散筒串方向进行，在本章中我们使用了 8 个离散叠加方向。

剂量的定义为单位质量中接收到的能量，在直角坐标系统中的 \mathbf{x} 点处剂量为：

$$D(\mathbf{x}) = \frac{E_{tot}(\mathbf{p})}{\rho_w(\mathbf{x}) \det(J(M(\mathbf{x})))} \quad (2.16)$$

其中点 $\mathbf{p}=M(\mathbf{x})$ 是点 \mathbf{x} 在改进后球坐标系统中的位置, J 表示两坐标系统之间映射关系 M 的雅克比行列式, 雅克比行列式值 $\det(J(M(\mathbf{x})))$ 是用来计算由直角坐标系统变换到改进后的球坐标系统所引起的体素体积变化。利用插值算法计算 $E_{tot}(\mathbf{p})$, 以减少由两坐标系统相互转换时带来的误差。

2.3.3 使用 FFTC 技术加速剂量计算

为了减少剂量计算所需时间, 本章中使用 FFTC 技术来加快剂量计算速度。Zhu 等人在直角坐标系统中使用了 FFTC 技术对剂量计算方法进行加速^[7]。但是这种方法在使用 FFTC 技术时认为由射线源发出的射线是相互平行的, 没有考虑射束的发散传播。在球坐标系统中, 使用不同的笔形束函数模型和非均匀介质中剂量分布的校正方法都会影响到 FFTC 技术的使用。在 AAA 算法^[6]中, 深度方向剂量是以射束中心轴与体模表层的入射点为参考点, 移动射束中心轴与球壳的交点来进行校正的。按这种校正方法, 原来在同一深度的球壳上的点经过对深度方向分量的校正后, 移动到不同深度的球壳上。而且其侧向分量中指数函数中的校正因子包含有深度信息, 且很难与侧向信息分离。因此, 在考虑到深度方向分量校正时, AAA 算法是不能使用 FFTC 技术进行加速的。在本章中, 深度方向分量是以射束中心轴与球壳的交点为参考点, 移动射束中心轴与体模的入射点和改变该射束的 DSS 来实现的。这种方法避免了移动射束中心轴和球壳交点所在球壳上的所有剂量计算点的位置, 同时也不需要在侧向分量中乘以包含深度信息的校正因子。因此, 新的笔形束剂量计算方法可以直接在每一层球壳上利用 FFTC 技术进行加速。

为了能在新的笔形束剂量计算方法中直接使用 FFTC 技术进行加速, 公式 2.11 中的包含有深度深度信息的分量 $I_{DSS}(\mathbf{p})c_{DSS, i}(\mathbf{p})$ 必须和包含有侧向信息的分量分离开来。在深度为 \mathbf{p}_z 的球壳上的剂量分布, 可以分两步计算得到: 首先分别计算计算包含深度信息的分量和包含侧向信息的分量并对两部分分量求傅立叶变换, 然后对两部分分量在频域中的乘积求逆傅立叶变化得到该球壳上的剂量分布。由于笔形束核是由 6 个指数函数的加权和构成, 总的傅立叶变换也应该是 6 个傅立叶变换乘积和的形式, 其公式如下:

$$E_{tot}(\mathbf{p}) = F^{-1} \left\{ \sum_{i=1}^{N=6} F \{ I_{DSS}(\mathbf{p}) c_{DSS, i}(\mathbf{p}) \} F \left\{ \frac{1}{\mu_i} e^{-\mu_i \sqrt{\mathbf{p}_x^2 + \mathbf{p}_y^2}} \right\} \right\} \quad (2.17)$$

2.4 实验材料

在 Tillikainen 等人^[6]提出的 5 种假设的体模中检验新方法的有效性, 这 5 种假设体模为: 1) 水体模; 2) 肺挡板体模, 一个厚 15cm 的低密度物质($\rho_w=0.3\text{g/cm}^3$)挡板置于水体模中距离表层 5cm 的位置处; 3) 骨挡板体模, 一个厚 5cm 的高密度物质($\rho_w=1.8\text{g/cm}^3$)挡板置于水体模中距离表层 5cm 的位置处; 4) 肺阻块体模, 一个厚 10cm 的低密度物质($\rho_w=0.3\text{g/cm}^3$)挡块置于水体模中距离表层 5cm, 且距离射野中心轴 2cm 的位置处; 5) 骨阻块体模, 一个厚 5cm 的高密度物质($\rho_w=1.8\text{g/cm}^3$)挡块置于水体模中距离表层 5cm, 且距离射野中心轴 2cm 的位置处。图 2-4 为这实验中使用的 5 种假设体模的结构。分别计算不同射野大小下的深度剂量曲线和侧向剂量曲线来评价新算法的有效性。对于水体模, 我们计算了射野大小分别为 30×30 , 50×50 , 100×100 和 $200 \times 200\text{mm}^2$ 的深度剂量曲线和射野为 $200 \times 200\text{mm}^2$ 的情况下, 深度分别在 50, 100 和 200mm 处的侧向剂量曲线。在肺挡板体模和骨挡板体模中分别计算了这 4 种不同射野情况下的深度剂量曲线。而在肺阻块体模计算了射野面积为 $100 \times 100\text{mm}^2$ 的情况下, 深度分别在 100 和 160mm 处的侧向剂量曲线; 在骨阻块体模中, 使用相同的射野, 计算深度分别在 75 和 110mm 处的侧向剂量曲线。

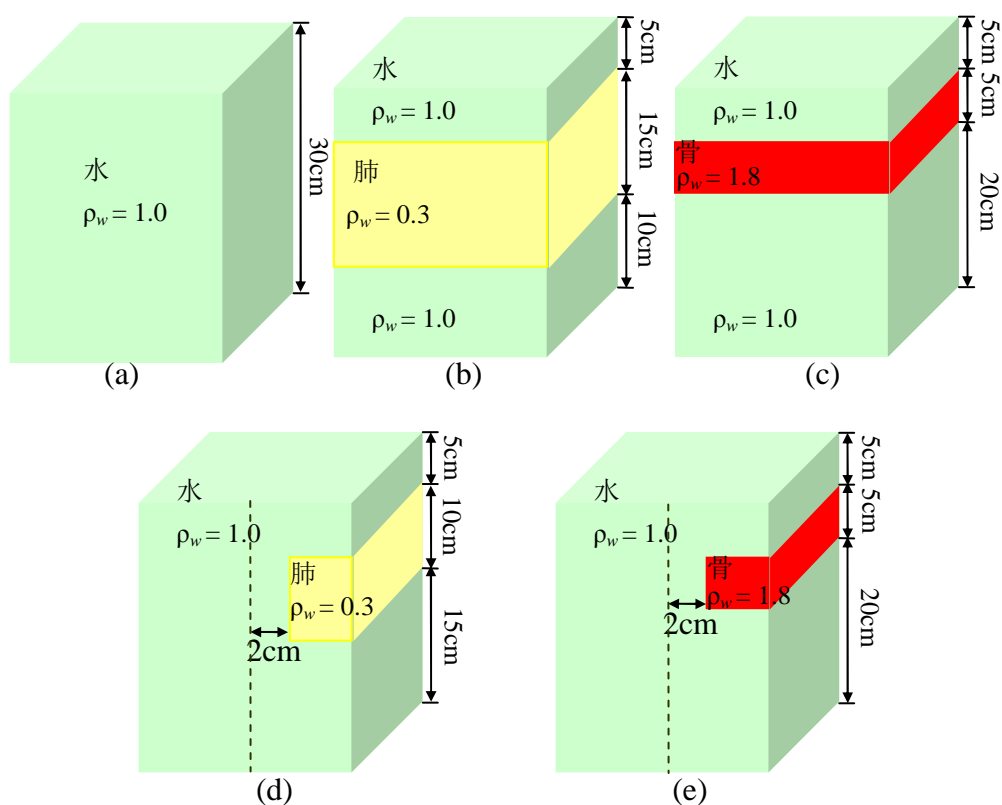


图 2-4 实验中使用的 5 种假设体模：（a）水体模；（b）肺挡板体模；（c）骨挡板体模；（d）肺阻块体模；（e）骨阻块体模。

在实验过程中，我们对不同的射野大小使用相同的源皮距 1000mm。在本章中，我们主要研究了体模内的剂量计算问题。因此，在实验过程中我们只针对简单射线源(点源)发出的原射线进行剂量计算，而没有考虑模拟真实治疗头时出现的射线经过均整器和准直器由于碰撞产生的散射以及射线在空气中电离等问题。本章中使用的点源为 6 MeV 多能光子源，其能谱来自于 EGS4 的能谱库(文件 mohan6.spectrum^[25])。本章提出的算法还与 MC 方法和 AAA 算法进行了比较。使用 MC 方法的剂量计算是利用开源软件 DOSXYZnrc^[26, 27]实现的。使用 DOSXYZnrc 软件进行剂量计算所使用的基本参数如下：体素边长=0.5cm，ECUT (electron/positron minimum transport energy) =0.700MeV 和 PCUT (photon minimum transport energy) =0.01MeV。在每个实验中，需要使用 MC 方法模拟 10 亿个光粒子在体模内的剂量分布，以保证剂量计算结果中的标准差在 0.1% 内。本章中所使用的 AAA 算法是根据 Tillikainen 等人^[6]提出的方法自己编程实现的。

2.5 实验结果

2.5.1 水体模

我们使用叠加剂量计算方法和 MC 方法分别在水体模中计算射野大小为 30×30, 50×50, 100×100 和 200×200mm² 的深度剂量曲线和射野为 200×200mm² 情况下，深度为体模表层下 50, 100 和 200mm 位置处的侧向剂量曲线，如图 2-5 所示。图 2-5(a)为两种算法在不同射野大小下的深度剂量曲线，其中叠加剂量计算方法的剂量结果用不同颜色的虚线表示，MC 方法的剂量结果使用不同颜色且不同形状的符号表示。在图 2-5 中，“Sup”表示叠加剂量计算方法得到的结果，“MC”表示 MC 方法计算得到的结果，“Fs”表示射野大小。在剂量计算方法中，MC 方法被认为是“金标准”，其剂量计算结果通常被用来代替真实的剂量分布。与 MC 方法的结果相比，在这 4 种射野中使用叠加剂量计算方法得到的深度剂量曲线的平均绝对误差依次为 0.32%，0.31%，0.34%和 0.84%。图 2-5(b)为两种算法在射野为 200×200mm²

的情况下, 体模内不同深度处的侧向剂量分布, “D” 表示侧向剂量曲线所在深度位置。使用叠加剂量计算方法得到的这 3 种不同深度下侧向剂量曲线的平均绝对误差依次为 1.64%, 1.52% 和 0.85%。图 2-5(c) 为图 2-5(b) 中深度为 200mm 的侧向剂量曲线, 其中 “Sup-MC” 表示两组侧向剂量曲线差值的绝对值。

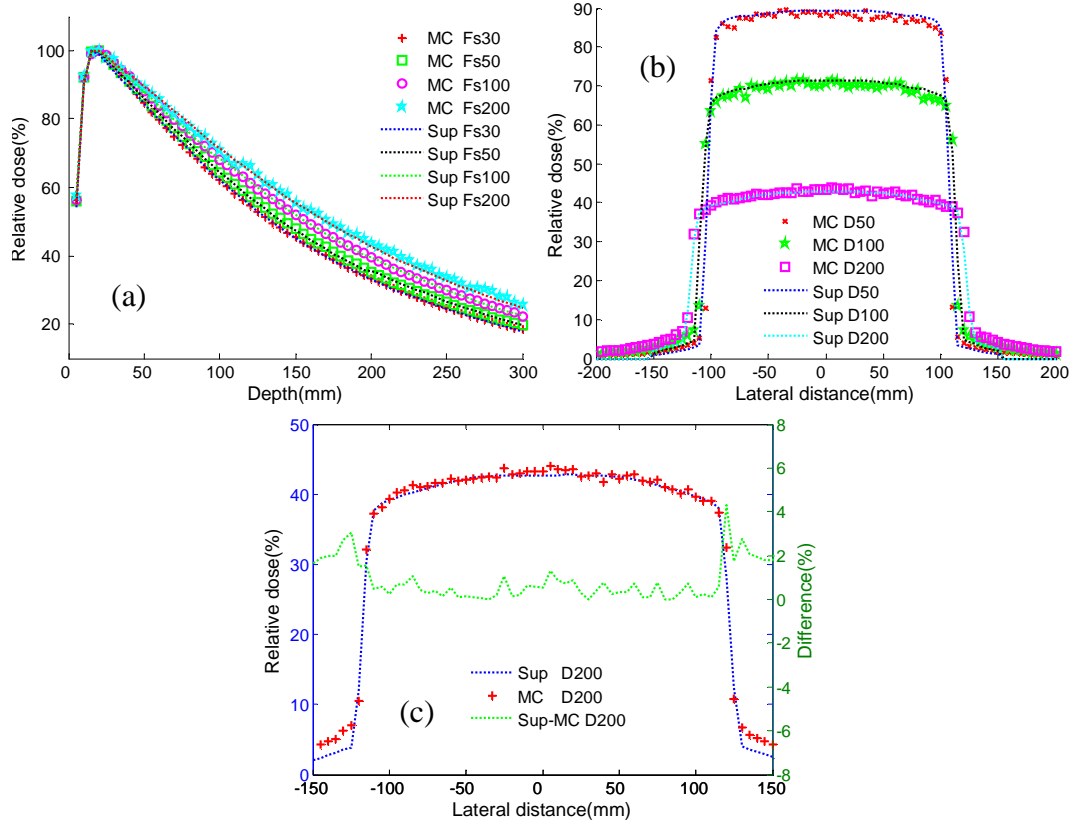
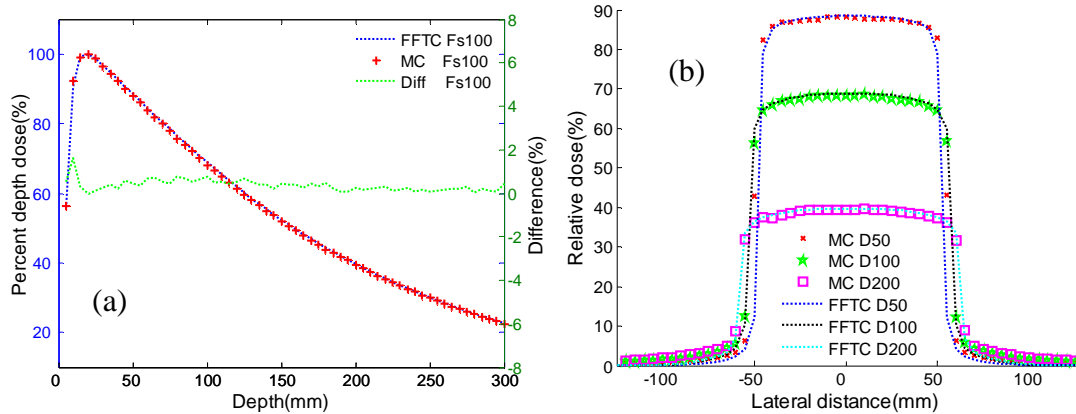


图 2-5 使用叠加剂量计算方法(Sup)和 MC 方法(MC)在水体模中剂量计算的结果: (a) 深度剂量曲线; (b) 不同深度下的侧向剂量曲线; (c) 深度为 200mm 的侧向剂量曲线。

同样, 我们在水体模中比较了使用 FFTC 技术加速的剂量计算方法和 MC 方法在射野为 $100 \times 100 \text{ mm}^2$ 情况下的剂量分布, 如图 2-6 所示。“FFTC” 表示使用 FFTC 技术加速的剂量计算方法得到的结果。图 2-6(a) 为两种算法的深度剂量曲线, 使用 FFTC 技术加速的剂量计算方法所得到的深度剂量曲线的平均绝对误差为 0.55%。图 2-6(b) 为两种算法在深度为体模表层下 50, 100 和 200mm 位置处的侧向剂量曲线, 使用 FFTC 技术加速的剂量计算方法所得到的侧向剂量曲线的平均绝对误差依次为 1.76%, 0.58% 和 1.07%。图 2-6(c) 为图 2-6(b) 中深度为 200mm 处的侧向剂量曲线, 其中 “FFTC-MC” 表示两组侧向剂量曲线差值的绝对值。



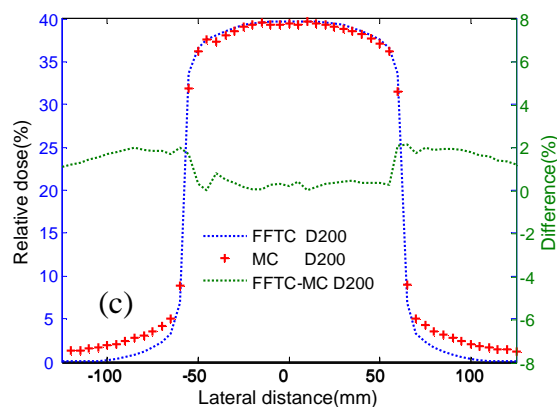


图 2-6 使用 FFTC 技术加速的剂量计算方法 (FFTC) 和 MC 方法 (MC) 在水体模中剂量计算的结果：(a) 深度剂量曲线；(b) 不同深度下的侧向剂量曲线；(c) 深度为 200mm 的侧向剂量曲线。

2.5.2 肺挡块体模

在肺挡块体模中，我们使用叠加剂量计算方法和 MC 方法分别计算射野大小为 30×30 ， 50×50 ， 100×100 和 $200 \times 200 \text{ mm}^2$ 情况下的深度剂量曲线。图 2-7 为不同射野大小下的深度剂量曲线，蓝色虚线为叠加剂量计算方法的计算结果，红色的十字符号为 MC 方法的计算结果，绿色虚线为两种方法计算结果差的绝对值。叠加剂量计算方法能精确地模拟光子射线在体模内的衰减过程，如在这 4 种不同射野大小的情况下，剂量计算结果的平均绝对误差依次为：0.72%，0.73%，0.83% 和 1.30%。叠加算法不能精确地模拟电离不平衡区域的剂量分布，从而导致了在两种介质的界面或射野边缘的电离不平衡区域内，剂量计算结果的误差比较大，在这 4 种射野下的电离不平衡区域内，最大的绝对误差分别为：3.76%，3.15%，2.73% 和 4.45%。

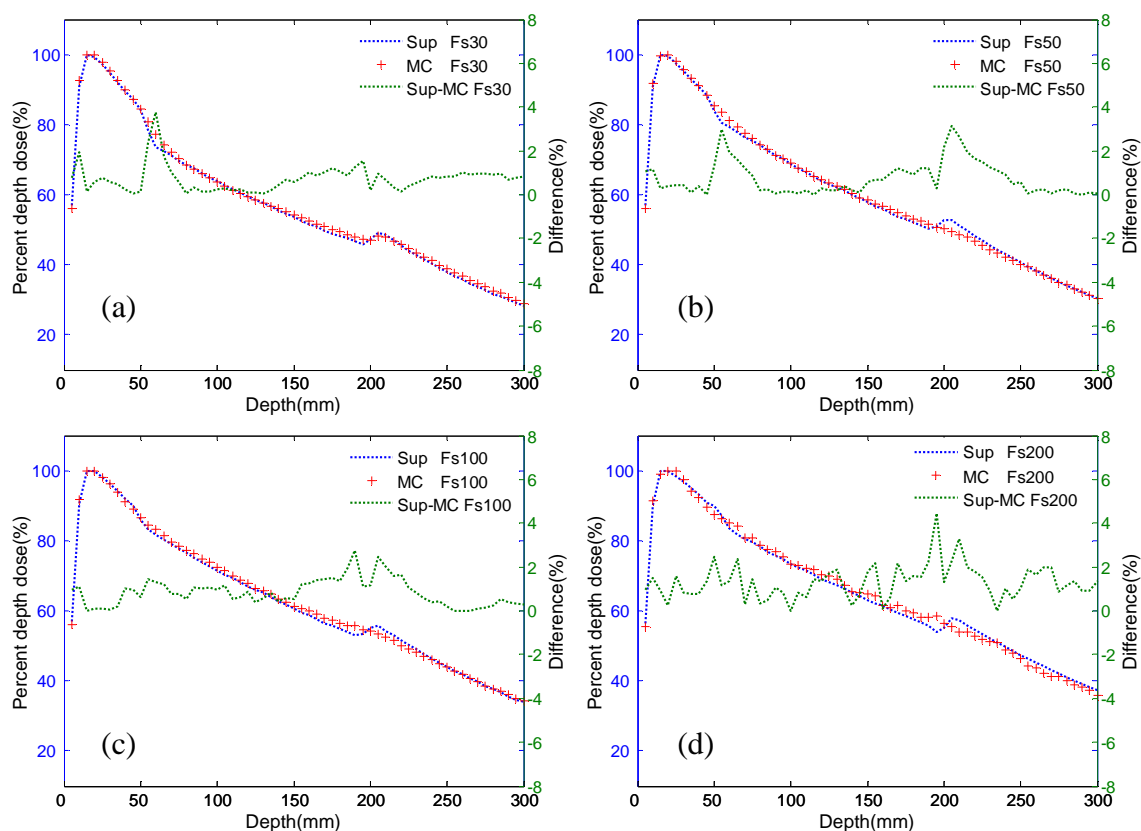


图 2-7 使用叠加剂量计算方法 (Sup) 和 MC 方法 (MC) 在肺挡块体模中剂量计算的结果。两种方法在不同射野大小情况下的深度剂量曲线：(a) $30 \times 30 \text{ mm}^2$ (Fs30)；(b) $50 \times 50 \text{ mm}^2$ (Fs50)；(c) $100 \times 100 \text{ mm}^2$ (Fs100)；(d) $200 \times 200 \text{ mm}^2$ (Fs200)。

2.5.3 骨挡块体模

在骨挡块体模中，我们使用叠加剂量计算方法和 MC 方法分别计算射野大小为 30×30 ， 50×50 ， 100×100 ，和 $200\times 200\text{mm}^2$ 情况下的深度剂量曲线，如图 2-8 所示。在这 4 种不同射野大小情况下，叠加剂量计算方法计算结果的平均绝对误差依次为：0.47%，0.83%，0.48% 和 0.56%。在两种介质的界面或射野边缘的电离不平衡，导致了叠加算法不能精确地模拟该区域的剂量分布，在这 4 种射野下的电离不平衡区域内，最大的绝对误差分别为：1.56%，2.18%，1.54% 和 2.55%。

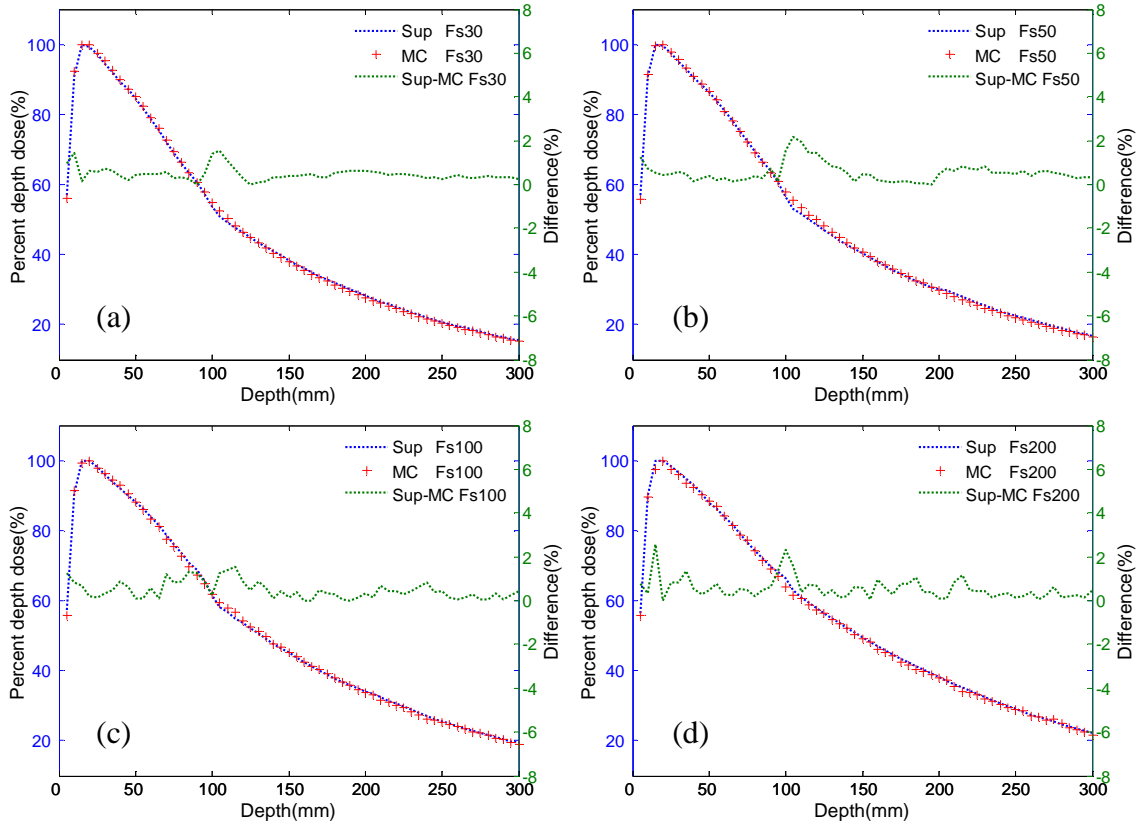


图 2-8 使用新的叠加剂量计算方法(Sup)和 MC 方法(MC)在骨挡块体模中剂量计算的结果。两种方法在不同射野大小情况下的深度剂量曲线：(a) $30\times 30\text{mm}^2$ (Fs30)；(b) $50\times 50\text{mm}^2$ (Fs50)；(c) $100\times 100\text{mm}^2$ (Fs100)；(d) $200\times 200\text{mm}^2$ (Fs200)。

2.5.4 肺阻块体模和骨阻块体模

在肺阻块体模和骨阻块体模中，我们使用叠加剂量计算方法和 MC 方法分别计算射野大小为 $100\times 100\text{mm}^2$ 情况下不同深度的侧向剂量曲线，如图 2-9 所示。叠加剂量计算方法能精确地模拟光子射线在这两种阻块体模内的输运过程。图 2-9(a)-(b)分别为肺阻块体模中距离体模表层 100mm 和 160mm 位置处的侧向剂量曲线，利用叠加剂量计算方法所得计算结果的平均绝对误差分别为 0.84% 和 0.72%。图 2-9(c)-(d)分别为骨阻块体模中距离体模表层 75mm 和 110mm 位置处的侧向剂量曲线，利用叠加剂量计算方法所得计算结果的平均绝对误差分别为 0.50% 和 0.60%。由于两种物质的界面或射野边缘存在电离不平衡的缘故，在这些区域内产生的最大剂量分别为误差：5.93%，4.18%，3.95% 和 3.97%，而在这些区域之外的绝对误差都在 1% 之内。

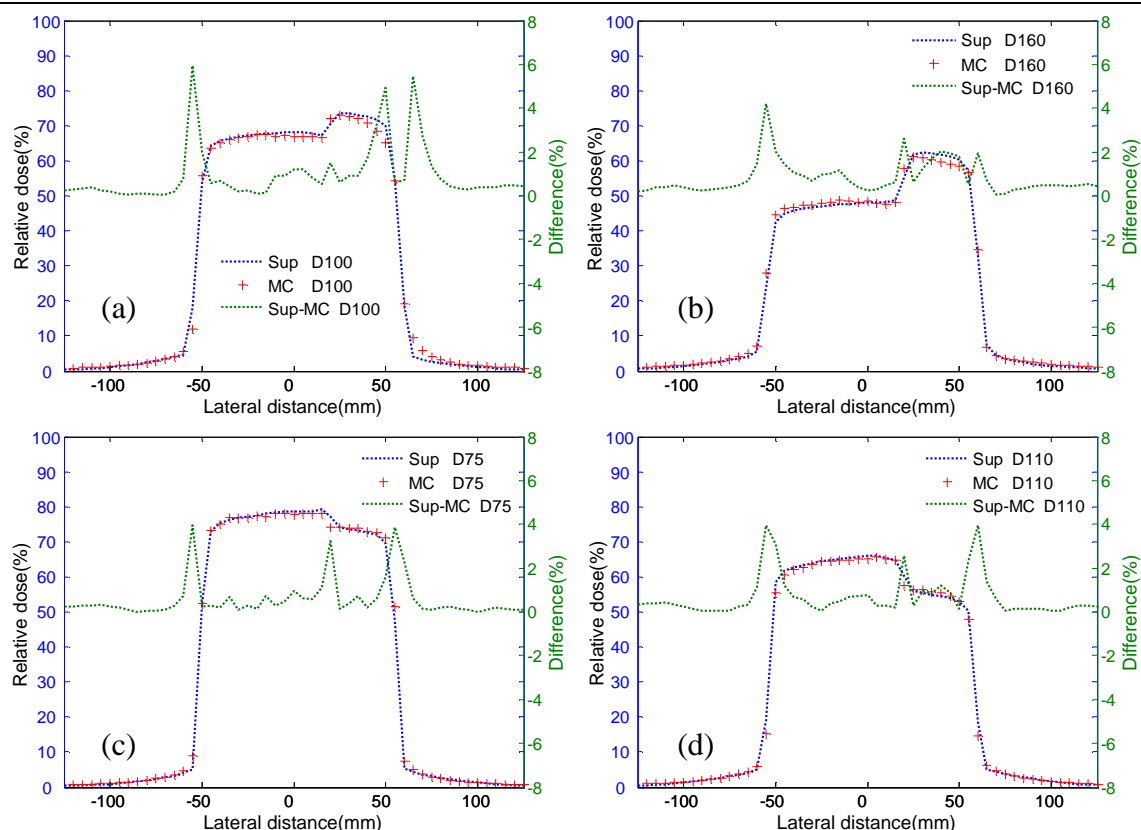


图 2-9 使用新的叠加剂量计算方法(Sup)和 MC 方法(MC)在肺阻块体模和骨阻块体模中剂量计算的结果。两种方法在不同射野大小为 $100 \times 100 \text{ mm}^2$ 情况下：(a) 肺阻块中 100mm 位置处的侧向剂量曲线；(b) 肺阻块中 160mm 位置处的侧向剂量曲线；(c) 骨阻块中 75mm 位置处的侧向剂量曲线；(d) 骨阻块中 110mm 位置处的侧向剂量曲线。

2.5.5 与 AAA 算法比较

本章提出的叠加剂量计算方法和 FFTC 技术加速后的剂量计算方法还与 AAA 算法在前面提到的 5 种体模中进行了比较。图 2-10 为这 3 种算法及 MC 方法在水体模、肺挡板体模以及骨挡板体模中分别计算射野大小为 $50 \times 50 \text{ mm}^2$ 和 $100 \times 100 \text{ mm}^2$ 情况下所得深度剂量曲线。图 2-11 是这 4 种算法在水体模、肺阻块体模以及骨阻块体模中计算射野大小为 $100 \times 100 \text{ mm}^2$ 情况下所得不同深度位置处的侧向剂量曲线。其中“AAA”表示 AAA 算法，“AAA-MC”表示 AAA 算法和 MC 方法计算结果差的绝对值。表 2.1 为在射野大小为 $100 \times 100 \text{ mm}^2$ 的情况下，本章提出的叠加剂量计算方法、FFTC 技术加速后的剂量计算方法以及 AAA 算法与 MC 方法在 5 种体模中计算结果的平均绝对误差值。实验结果表明，本章提出的叠加剂量计算方法精度要略高于 AAA 算法和 FFTC 加速后的剂量计算方法，更接近 MC 方法的计算结果。相对于 AAA 算法，本章提出的叠加剂量计算方法在肺挡板体模中的平均绝对误差依次减少为 0.77% 和 0.76%，如图 2-10(c)-(d)；在骨挡板体模中依次减少为 0.18% 和 0.35%，如图 2-10(e)-(f)。相对于 MC 方法，其余 3 种方法在射野大小为 $100 \times 100 \text{ mm}^2$ 情况下的详细比较见表 2.1。

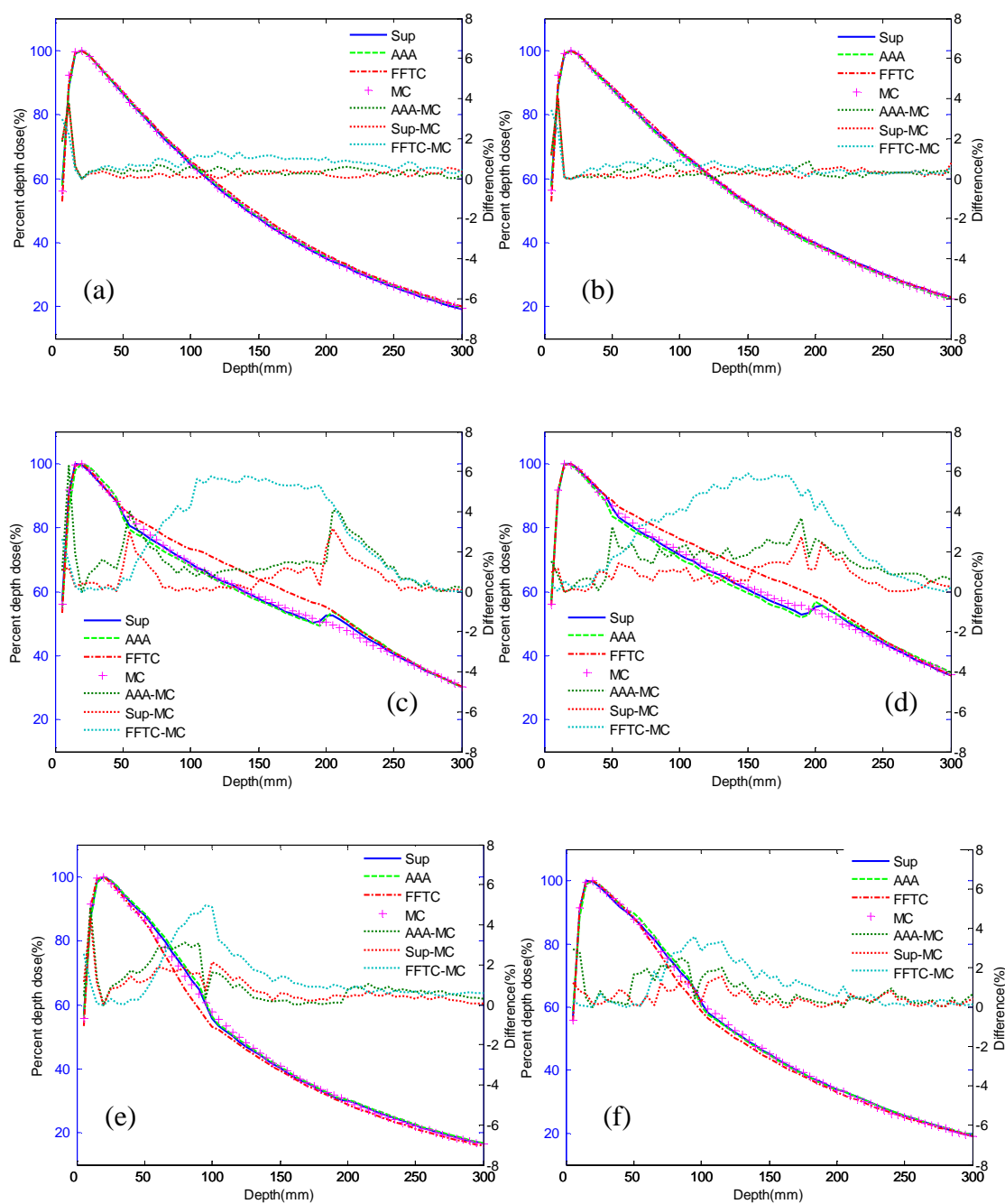


图 2-10 使用新的叠加剂量计算方法(Sup)、FFTC 技术加速后的剂量计算方法(FFTC)、AAA 算法(AAA)和 MC 方法(MC)在水体模、肺挡块体模以及骨挡块体模中剂量计算所得深度剂量曲线：(a) 水体模中射野为 $50 \times 50 \text{mm}^2$ 的深度剂量曲线；(b) 水体模中射野为 $100 \times 100 \text{mm}^2$ 的深度剂量曲线；(c) 肺挡板体模中射野为 $50 \times 50 \text{mm}^2$ 的深度剂量曲线；(d) 肺挡板体模中射野为 $100 \times 100 \text{mm}^2$ 的深度剂量曲线；(e) 骨挡板体模中射野为 $50 \times 50 \text{mm}^2$ 的深度剂量曲线；(f) 骨挡板体模中射野为 $100 \times 100 \text{mm}^2$ 的深度剂量曲线。

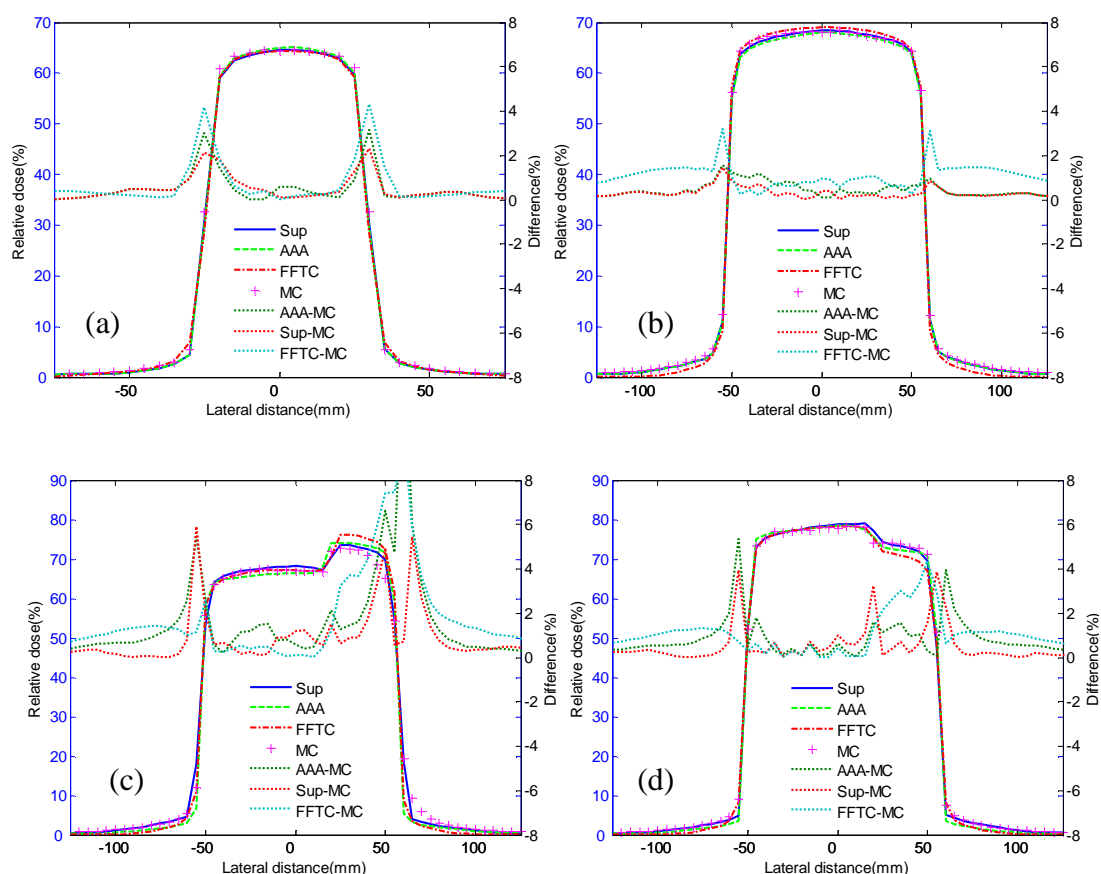


图 2-11 使用新的叠加剂量计算方法(Sup)、FFTC 技术加速后的剂量计算方法(FFTC)、AAA 算法(AAA)和 MC 方法(MC)在水体模、肺挡块体模以及骨挡块体模中剂量计算所得侧向剂量曲线：(a) 水体模中射野为 $50 \times 50 \text{ mm}^2$ 深度为 100mm 位置处的侧向剂量曲线；(b) 水体模中射野为 $100 \times 100 \text{ mm}^2$ 深度为 100mm 位置处的侧向剂量曲线；(c) 肺阻块体模中射野为 $100 \times 100 \text{ mm}^2$ 深度为 100mm 位置处的侧向剂量曲线；(d) 骨阻块体模中射野为 $100 \times 100 \text{ mm}^2$ 深度为 75mm 位置处的侧向剂量曲线。

表 2.1 在射野大小为 $100 \times 100 \text{ mm}^2$ 的情况下，本章提出的叠加剂量计算方法、FFTC 技术加速后的剂量计算方法以及 AAA 算法与 MC 方法在 5 种体模中计算结果的平均绝对误差值。

	深度剂量曲线			侧向剂量曲线		
	水体模	肺挡板体模	骨挡板体模	水体模	肺阻块体模	骨阻块体模
AAA	0.37%	1.59%	0.83%	0.29%	1.47%	0.77%
Sup	0.34%	0.83%	0.48%	0.22%	0.84%	0.50%
FFTC	0.55%	2.89%	1.05%	0.58%	1.67%	0.98%

2.5.6 计算时间

本章提出的叠加剂量计算方法、FFTC 技术加速后的剂量计算方法以、AAA 算法和 MC 方法在射野大小分别为 $100 \times 100 \text{ mm}^2$ 和 $200 \times 200 \text{ mm}^2$ 的情况下所需计算时间如表 2.2 所示。所有的算法都是在同一台电脑上实现的(双核 Intel(R) Core(TM) 2 CPU 6600, 2GB 内存)。本章提出的叠加剂量计算方法在这两种射野下所需计算时间分别为 159.8s 和 182.3s。与 AAA 算法相比，本章提出的叠加剂量计算方法所需计算时间较少，没有增加计算复杂度。由于利用 FFTC 技术加速的剂量计算方法忽略了体模内侧向非均匀介质的分布，因此其剂量计算精度较低。但是，使用 FFTC 技术加速后，其计算速度非常快。在 $64 \times 64 \times 60$ 个点的体模中，完成这

两种射野下的剂量计算只需 4.0s 和 4.1s。

表 2.2 本章提出的叠加剂量计算方法、FFTC 技术加速后的剂量计算方法、AAA 算法和 MC 方法在射野大小分别为 $100\times 100\text{mm}^2$ 和 $200\times 200\text{mm}^2$ 情况下所需的计算时间。

	$100\times 100\text{mm}^2$	$200\times 200\text{mm}^2$
FFTC	4.0s	4.1s
Sup	159.8s	182.3s
AAA	170.8s	184.4s
MC	12h	14h

2.6 讨论

在本章中，主要针对校正射线发散所使用的坐标系统和 AAA 算法在非均匀介质中的剂量校正以及使用 FFTC 技术加快剂量计算中所存在的问题，提出了一种新的笔形束剂量计算模型。在新的笔形束模型中，我们实现了叠加剂量计算方法并对该方法使用 FFTC 技术加速。实验结果表明，新提出的剂量计算方法精度要略高于 AAA 算法，在射野大小为 $100\times 100\text{mm}^2$ 情况下，其平均绝对误差小于 1%。但是，叠加算法不能精确模拟电离不平衡区域的剂量分布，从而导致了在两种介质的界面或射野边缘的剂量计算结果误差较大(高达 6%)。使用 FFTC 技术加速后的剂量计算方法精度较低(如表 2.1 所示)，但是其速度相对于叠加算法提升了将近 40 倍(如表 2.2 所示)。

已有在球坐标系统中使用笔形束剂量算法来校正射线发散而不增加计算量的方法，这类方法已在一些临床治疗系统中使用^[4, 6, 10, 11]。由射线源发出的射线，除射束中心轴处的射线垂直于体模表层入射，其余射线都从体模表层斜倾入射。射线倾斜入射时，其在均匀物质中的能量分布不再呈轴对称分布，而且在不同的入射角度下的射束所对应的笔形束核相互之间也有差异。AAA 算法^[4, 6]在剂量计算过程中考虑到了不同入射角度下笔形束核的之间差异，因而模拟了不同入射角度下的笔形束核。而本章提出的笔形束模型，主要考虑了在非均匀介质中的剂量校正，使用不同的校正方法来避免由校正算法本身引入的误差。新的校正方法以射束中心轴与球壳的交点为参考点，通过移动射束中心轴与体模表层的交点和改变该射束的 DSS 来实现的。新方法需要模拟不同 DSS 下的笔形束核。本章使用 MC 方法模拟平行射束在零射野下的射束能量分布来得到笔形束核。该笔形束核与 DSS 大小无关。在改进后的球壳坐标系统中，不同 DSS 的笔形束核根据公式 2.15 对得到笔形束核进行计算得到。由于对不同入射角度下所得笔形束核模拟不同 DSS 的笔形束核，需要存储大量参数值，改进的笔形束模型没有考虑不同入射角度下射束所对应的笔形束核之间的差异。

为了提高剂量计算速度，在改进后的球坐标系统中使用 FFTC 技术加速剂量计算。在本章中提出的笔形束剂量计算模型中，深度方向剂量的校正以射束中心轴和球壳的交点为参考点。与 AAA 算法相比，在深度方向剂量校正后，原来在同一球壳上的剂量计算点的位置没有改变。因此，在不考虑侧向介质的非均匀性，可以直接在每层球壳上使用 FFTC 技术进行加速。加速后的算法大大地提高了剂量计算速度。由于使用 FFTC 技术加速，在剂量计算过程中不能精确地考虑侧向非均匀介质中的剂量校正。相比叠加算法，FFTC 技术加速后算法的精度较低。在 IMRT 方案优化中需要在尽可能短的时间内完成优化。已有学者在方案优化的迭代过程中是用混合剂量计算方法以减少方案优化所需时间^[28, 29]。在大部分的迭代过程中使用 FFTC 技术加速的剂量算法。在优化过程的开始和迭代一定次数后，使用 MC 方法或基于点核的剂量计算方法精确地计算剂量分布，来校正由 FFTC 技术加速后算法所得低精度的剂量计算结果。

2.7 本章小结

在本章中，我们提出了一种新的笔形束剂量计算模型来减少由球壳坐标系统中的采样方

式及在非均匀介质中的剂量校正所带来的误差，并在该模型上使用 FFTC 技术进行加速。新的笔形束模型分别使用叠加方法和 FFTC 进行剂量计算。相比 AAA 算法，叠加剂量计算方法在增加剂量计算精度的同时，没有增加剂量计算的时间。在不考虑侧向非均匀介质分布的情况下，可以在每一层球壳上直接使用 FFTC 技术对新的笔形束算法进行加速。

射线倾斜入射时，其在均匀介质中的剂量分布不在为轴对称分布。在本章中，不同 DSS 的笔形束核都是由一个轴对称的笔形束核计算得到。因此，直接使用不同 DSS 的笔形束核计算射线倾斜入射时的剂量分布，会引入一些小的误差。在将来的工作中，考虑不同入射角度下不同 DSS 的笔形束核，可以进一步增加笔形束剂量计算方法的精度。在本章中，新算法只计算了 6 MeV 多能谱点源下的剂量分布，而不是计算真实治疗头下的剂量分布。剂量计算的精度很大程度上与精确地模拟真实的治疗头有关。将射线源假设为一个点源和使用简单的多能谱源都会大大地减小剂量计算的精度。在将来的工作中，我们将模拟真实的治疗头，将治疗头模拟成多源模型，分别表示原射线、散射线、电离污染等，并在真实的病人 CT 数据中计算剂量分布。笔形束模型只是一种近似模拟射线在体模内剂量分布的剂量计算模型。相对于基于点核模型的剂量计算方法和 MC 方法，笔形束模型由于自身结构的局限性，其剂量计算的精度还是比较低的。

2.8 参考文献

- [1] Siebers JV, Tong S, Lauterbach M, et al. Acceleration of dose calculations for intensity-modulated radiotherapy [J]. Med Phys, 2001, 28(6): 903-910.
- [2] Ahnesjö A, Saxner M, Trepp A. A pencil beam model for photon dose calculation [J]. Med Phys, 1992, 19(2):263-73.
- [3] Storchi PRM, van Battum LJ, Woudstra E. Calculation of a pencil beam kernel from measured photon beam data [J]. Phys Med Biol, 1999, 44(12): 2917-2928.
- [4] Ulmer W, Pyyry J, Kaissl W. A 3D photon superposition/convolution algorithm and its foundation on results of Monte Carlo calculations [J]. Phys Med Biol, 2005, 50(8):1767-1790.
- [5] Jeleń U, Alber M. A finite size pencil beam algorithm for IMRT dose optimization: density corrections [J]. Phys Med Biol, 2007, 52(3):617-633.
- [6] Tillikainen L, Helminen H, Torsti T, et al. A 3D pencil-beam-based superposition algorithm for photon dose calculation in heterogeneous media [J]. Phys Med Biol, 2008, 53(14):3821-3839.
- [7] Zhu Y, Boyer A. X-ray dose computations in heterogeneous media using 3-dimensional FFT convolution [J]. Phys Med Biol, 1990, 35(3):351-368.
- [8] Erdi AK, Yorke ED, Loew MH, et al. Use of the fast Hartley transform for three-dimensional dose calculation in radionuclide therapy [J]. Med Phys, 1998, 25(11):2226-2233.
- [9] Battista JJ, Sharpe MB. True three-dimensional dose computations for megavoltage x-ray therapy: a role for the superposition principle [J]. Australas Phys Eng Sci Med, 1992, 15(4):159-178.
- [10] Lu W, Chen M. Fluence-convolution broad-beam (FCBB) dose calculation [J]. Phys Med Biol, 2010, 55(23):7211-7229.
- [11] Lu W. A non-voxel-based broad-beam (NVBB) frame work for IMRT treatment planning [J]. Phys Med Biol, 2010, 55(23):7175-7210.
- [12] Knöös T, Wieslander E, Cozzi L, et al. Comparison of dose calculation algorithms for treatment planning in external photon beam therapy for clinical situation [J]. Phys Med Biol, 2006, 51(22):5785-5807.
- [13] Lillicrap S C, Wilson P, Boag J W. Dose distributions in high energy electron beams: Production of broad beam distributions from narrow beam data [J]. Phys Med Biol, 1987, 20(1):30-38.

- [14] Boyer AL, Zhu YP, Wang L, et al. Fast Fourier transform convolution calculations of x-ray isodose distributions in homogeneous media [J]. *Med Phys*, 1989, 16(2):248-253.
- [15] Papanikolaou N, Mackie TR, Meger-Wells C, et al. Investigation of the convolution method for polyenergetic spectra [J]. *Med Phys*, 1993, 20(5):1327-1336.
- [16] Sharpe MB, Battista JJ. Dose calculations using convolution and superposition principles: the orientation of dose spread kernels in divergent x-ray beams [J]. *Med Phys*, 1993, 20(6):1685-1694.
- [17] Liu HH, Mackie TR, McCullough EC. Correcting kernel tilting and hardening in convolution/superposition dose calculations for clinical divergent and polychromatic photon beams [J]. *Med Phys*, 1997, 24(11):1729-1741.
- [18] Ahnesjö A, Aspradakis MM. Dose calculations for external photon beams in radiotherapy [J]. *Phys Med Biol*, 1999, 44(11):R99-R155.
- [19] Seco J, Evans PM. Assessing the effect of electron density in photon dose calculations [J]. *Med Phys*, 2006, 33(2):540-552.
- [20] McGary JE, Boyer AL. An interactive, parallel, three-dimensional fast Fourier transform convolution dose calculation using a supercomputer [J]. *Med Phys*, 1997, 24(4):519-522.
- [21] Ahnesjö A. A Collapsed cone convolution of radiant energy for photon dose calculation in heterogeneous media [J]. *Med Phys*, 1989, 16(4):577-592.
- [22] Aspradakis MM, Redpath AT. A technique for the fast calculation of three-dimensional photon dose distributions using the superposition model [J]. *Phys Med Biol*, 1997, 42(8):1475-1489.
- [23] Kawrakow I. Accurate condensed history Monte Carlo simulation of electron transport I EGSnrc, the new EGS4 version [J]. *Med Phys*, 2000, 27(3): 485-498.
- [24] C. Scholz, C. Schulze, U. Oelfke, et al. Development and clinical application of a fast superposition algorithm in radiation therapy [J]. *Radiother Oncol*, 2003, 69(1):79-90.
- [25] Mohan R, Chui CS, Lidofsky L. Energy and angular distributions of photons from medical accelerators [J]. *Med Phys*, 1985, 12(2):592-597.
- [26] Rogers DWO, Walters B, Kawrakow I. BEAMnrc users manual NRC report [M]. Ottawa, Canada: National Research Council of Canada; 2005.
- [27] Walters B, Kawrakow I, Rogers DWO: DOSXYZnrc users manual NRC report [M]. Ottawa, Canada: National Research Council of Canada; 2005.
- [28] Jeffrey VS, Marc L, Tong S, et al. Reducing dose calculation time for accurate iterative IMRT planning [J]. *Med Phys*, 2002, 29(2):231-237.
- [29] Jeffrey VS, Kawrakow I, Ramakrishnan V. Performance of a hybrid MC dose algorithm for IMRT optimization dose evaluation [J]. *Med Phys*, 2007, 34(7):2853-2863.

第三章 基于生物准则的 IMRT 方案优化算法的研究

3.1 引言

当前临床中使用最为广泛的是多叶准直器(multileaf collimator, MLC) 调强技术。MLC 调强技术主要有静态调强和动态调强两种模式。传统的 MLC 静态调强实现方法主要有：两步法和直接子野优化法两种。本章中主要考虑两步法中的注量图优化(fluence map optimization, FMO)问题。FMO 问题是一个寻找最优注量图分布的逆向计划过程，通过在解空间中寻找一组最优或可行的解，从而使得尽可能多的高剂量沉积于靶区内，同时尽可能避免过多的剂量照射靶区周围的关键组织和正常组织。FMO 模型的目标函数通常是由一组相互冲突的治疗方案评价准则构成，这些准则用来反映优化结果是否满足治疗目的或限制。

治疗方案的评价准则可以由物理准则或生物准则构成。物理准则通常被定义为处方剂量分布和当前剂量分布之差的二次型函数或者是剂量-体积(dose-volume, DV)限制条件^[1]。物理准则的放射生物效应的合理性是值得质疑的。对于同样剂量的偏移，高于处方剂量和低于处方剂量产生的生物效应是不同的。因此，物理准则不能精确地描述肿瘤或正常组织对放射剂量的非线性反应，特别是在非均匀剂量分布的情况下。而通过综合地考虑组织性能和潜在的放射生物反应，生物准则可以合理地预测肿瘤或正常组织在非均匀剂量分布下的生物效应。生物目标函数已成为一种替代物理目标函数的选择^[2]。把肿瘤控制率(tumor control probability, TCP)和正常组织并发症发生概率(normal tissue complication probability, NTCP)应用到逆向计划中，逆向计划会找到一个更好地平衡肿瘤控制和正常组织并发症的最优解。因此，使用基于生物准则的目标函数可以进一步改善对患者的治疗效果。

然而，生物准则通常是关于剂量分布的 S 形函数，如 TCP 和 NTCP。因此，生物准则是关于剂量分布的非线性非凸函数。直接在目标函数中使用这些非线性且非凸的函数，使得求解 FMO 问题变得更加困难，特别是在自变量（射束）比较多的情况下。在求解大规模的非凸函数的优化问题中，通常会遇到如下困难^[3]：1) 目标函数在局部最优点收敛；2) 可能不存在可行解；3) 优化算法所选择的终止条件可能过于武断；4) 优化问题可能被降级；5) 目标函数的收敛率比较低；6) 数值问题可能导致优化提前停止或无限制的执行下去。幸运的是，在求解最小值的优化问题中，如果所有的目标函数和条件都是凸函数，前三个困难将会被克服。但是目标函数的收敛率和数值稳定性仍然是一个潜在的问题。如果目标函数和条件同时还是自和谐障碍函数(self-concordant barriers functions)，那么可以使用内点法(interior point methods)来克服收敛率较低和数值敏感的问题^[4]。因此，可以使用局部求解算法高效地解决一大类凸优化问题，如基于梯度的优化算法。在多目标 FMO 优化问题中，使用凸目标函数同时保证了 Pareto 有效边界是凸的^[5]，特别是使用分段线性近似技术估算 Pareto 有效边界，如 Sandwich 算法^[6, 7]。

为了克服直接使用生物准则使得求解优化问题变难的问题，Romeijn 等人在多目标 FMO 优化问题中建立了一个将非凸生物准则转换为等效凸生物准则的统一框架，而不改变 Pareto 有效解^[5]。Hoffmann 等人^[8]进一步扩展了 Romeijn 等人^[5]的工作，将反应剂量分次照射效应的 LQ(linear-quadratic)细胞存活模型转换为等效凸生物准则。尽管 Romeijn 等人^[5]和 Hoffmann 等人^[8]提出了这种将非凸生物准则转化为等效凸生物准则的方法，但是他们没有在实际治疗计划中检验这种方法的有效性。在本章中，我们提出一种利用等效凸生物准则构造目标函数的方法并使用基于梯度的优化算法解 FMO 问题。总目标函数表示为体模内各器官上所用于子目标函数和的形式。子目标函数由等效凸准则的值与处方量的差值和该差值的阶跃函数的乘积构成。在优化开始之前，使用经典笔形束算法计算单位能量射束在体模内的剂量分布并存储该剂量效应矩阵。体模内的剂量分布为剂量沉积矩阵和注量分布的乘积。在本章中，使用基于梯度的优化算法，L-BFGS (limited-memory Broyden-Fletcher-Goldfarb-Shanno)算法^[9]，解 FMO 问题，还需要给出各子目标函数的一阶导数。

本章首先介绍了使用生物准则进行方案优化的优点，以及将不同物理准则和生物准则转换为等效凸函数的方法。然后，提出了一种利用等效凸生物准则构造目标函数的方法并使用基于梯度的优化算法解 FMO 问题。最后，在前列腺肿瘤病人数据上验证了提出算法的有效性，并分析了生物准则所使用的不同参数对优化结果的影响。

3.2 优化准则

3.2.1 物理准则

现在广泛使用的物理准则是以实际剂量与处方剂量(靶目标)或耐受剂量(靶目标和正常组织)的二次偏差的形式给出。利用不同的权重系数来惩罚过高或过低的剂量分布。我们将物理准则分为：最小剂量准则，最大剂量准则，平均剂量准则和剂量体积直方图(dose volume histogram, DVH)准则，并按二次偏差的形式给出部分准则的计算模型。

3.2.1.1 最小剂量准则

最小剂量准则是用来描述在靶区内实际剂量小于处方剂量的放射生物反应：

$$f_{\min}(\mathbf{D}) = \frac{1}{N} \sum_{i=1}^N H(D_{\min} - \mathbf{D}_i) \cdot (D_{\min} - \mathbf{D}_i)^2 \quad (3.1)$$

其中 D_{\min} 为靶区处方剂量， $H(x)$ 为阶跃函数， N 为器官内的所有体素点， \mathbf{D}_i 为器官第 i 个体素的剂量值。

3.2.1.2 最大剂量准则

最大剂量准则是用来描述在靶区或正常组织内实际剂量大于耐受剂量的放射生物反应：

$$f_{\max}(\mathbf{D}) = \frac{1}{N} \sum_{i=1}^N H(\mathbf{D}_i - D_{\max}) \cdot (\mathbf{D}_i - D_{\max})^2 \quad (3.2)$$

其中 D_{\max} 为耐受剂量。

3.2.1.3 平均剂量准则

平均剂量准则即可描述靶区剂量分布的均匀程度，也是描述并行正常组织放射效应的一个重要临床参数，其公式如下：

$$f_{\text{mean}}(\mathbf{D}) = \frac{1}{N} \sum_{i=1}^N (\mathbf{D}_i - D_{\text{mean}})^2 \quad (3.3)$$

其中 D_{mean} 为靶区处方剂量或正常组织的平均耐受剂量。

3.2.1.4 DVH 准则

DVH 准则是一种常用的放射计划评估方法，既可以评价靶区剂量的均一性，也可以描述正常组织内剂量-体积的依赖程度。DVH 准则也被称为 DV 约束。DVH 准则定义为器官内不超过 $V_{\max}\%$ 的体积接受到的剂量超过 $D_{\max}^{[15]}$ 。DVH 准则可以被视为 DVH 曲线必须通过 (D_{\max}, V_{\max}) 点的一个约束条件。有时，需要在同一个器官上使用多个 DVH 约束来控制其剂量分布。由于 DVH 准则的定义与处方剂量或耐受剂量没有直接的联系，因此，不能直接表示为二次偏差的形式。

3.2.2 生物准则

3.2.2.1 EUD 准则

Niemierko^[16]最早提出了等效均一剂量(equivalent uniform dose, EUD)的概念，是通过描述与实际剂量照射靶目标产生相同的肿瘤控制率(杀灭同样数目的细胞)时所需的均匀剂量照

射靶目标而提出来的，其定义为：

$$EUD(\mathbf{D}) = -\frac{1}{\alpha} \ln\left(\frac{1}{N} \sum_{i=1}^N e^{-\alpha \mathbf{D}_i}\right) \quad (3.4)$$

其中 N 为靶目标的体素数目， \mathbf{D}_i 为在第 i 个体素内的受照剂量； α 为剂量体积效应因子。利用这种方法计算得到的 EUD 值，受细胞增殖速率、分次照射次数以及分次照射之间的时间间隔的影响。基于由 Kutcher 和 Burman^[17] 提出的离散模型，Niemierko^[18] 随后给出了广义 EUD (generalized EUD, gEUD) 的概念，即在靶区内的广义平均剂量分布：

$$gEUD(\mathbf{D}) = \left(\frac{1}{N} \sum_{i=1}^N \mathbf{D}_i^a \right)^{1/a} \quad (3.5)$$

其中 $-\infty \leq a \leq 0$ 为反应剂量体积效应的参数因子，用来描述靶区组织的剂量效应。当参数 $1 \leq a \leq +\infty$ 时，gEUD 的概念被扩展到了正常组织上。当 a 分别为 $-\infty$ 、 0 和 $+\infty$ 时，对应的 gEUD 值分别为该组织内剂量分布的最小值、均值和最大值。

(g)EUD 模型包括一个参数 a ，选取不同的 a 值体现不同组织在相同剂量下的不同放射生物效应。在并行器官中，正常组织内的功能亚单元(function subunit, FSU)是相互独立的，足够小区域内的 FSU 受到损坏不会使整个器官的功能失调，如肺，肾和肝。因此，并行器官对整体的剂量分布比较敏感，通常选取较小的 a 值来描述并行器官对整体剂量分布的放射生物效应。而在串行器官中，正常组织内的 FSU 是排成一列的，如果一个功能单元被破坏，整个器官会出现并发症，如脊髓，肠和视神经。因此，并行器官对剂量分布内的高剂量比较敏感，通常选取较大的 a 值来描述串行器官对剂量分布内高剂量的放射生物效应。

3.2.2.2 TCP 准则

剂量效应曲线的基本形状是 S 型。TCP 不仅是剂量的函数，而且特别依赖于细胞有氧状态、癌细胞血管生成量等因素。在同一个肿瘤的不同部位这些因素的作用是不同的，而且还可能随时间改变，甚至在很短的放射治疗过程中就会发生改变。在文献中已经提出了一些 TCP 模型^[19-30]。所有的模型都是基于相同的假设，即肿瘤是由一些相互独立的克隆源性细胞组成，且这些克隆源性细胞对照射剂量的反应是相互独立的。

1) Schultheiss 模型

Schultheiss 模型起初只用于评估正常组织的剂量效应^[19]，因为 TCP 和 NTCP 曲线均为 S 形，Schultheiss 逻辑模型也可被用来表示肿瘤的剂量效应：

$$TCP(D) = \frac{1}{1 + (D_{50}/D)^k} \quad (3.6)$$

式中 D_{50} 为 $TCP = 0.5$ 时的剂量值， k 为描述剂量效应曲线特征的一个参数。为了描述非均匀剂量分布下靶目标的放射生物效应，Niemierko 将式 3.6 中的剂量分布 D 用 EUD 来替代^[20]。

2) LQ-泊松模型

靶目标内所有的肿瘤克隆源性细胞都被杀灭的概率，通常被称为肿瘤控制率^[21]，即 TCP。靶区内的每一个肿瘤单元都得被根除以达到靶区内所有的肿瘤克隆源性细胞都被杀灭。因此 TCP 的值为靶区内所有肿瘤单元放射反应的乘积。用来描述靶区剩余的肿瘤克隆源性细胞以及细胞增殖行为的基于泊松近似分布的 TCP 公式可以表示为：

$$TCP(\mathbf{D}) = \prod_{i=1}^v \exp\left(-\frac{N_0 e^{\lambda(n-1)\Delta T}}{v} \cdot p(\mathbf{D}_i)\right) \quad (3.7)$$

其中 N_0 是靶区内所有的肿瘤克隆源性细胞数， $p(\mathbf{D}_i)$ 表示在靶区第 i 个体素内的克隆源性细胞在受到剂量 \mathbf{D}_i 照射后的存活率， n 是进行放射治疗总的分次照射次数， ΔT 为连续两次分次照

射之间的时间间隔, λ 为细胞净增殖速率。Stavreva 等人^[22]提出包含细胞增殖单击模型可以非常好地拟合肿瘤的放射生物效应。在这种情况下, TCP 公式变为:

$$TCP(\mathbf{D}) = \prod_{i=1}^v \exp\left(-\frac{N_0 e^{\lambda(n-1)\Delta T}}{v} \cdot e^{-\mathbf{D}_i / D_{lethal}}\right) \quad (3.8)$$

其中 D_{lethal} 为平均致死剂量, 描述靶区内细胞的耐辐射性。

在不考虑细胞增殖的情况下, Fowler^[23]提出了基于线性二次(linear-quadratic, LQ) 泊松模型的 TCP 公式:

$$TCP(\mathbf{D}) = \exp\left[-\frac{N_0}{N} \sum_{i=1}^N SF_{LQ}(\mathbf{D}_i)\right] \quad (3.9)$$

其中 $SF_{LQ}(\mathbf{D}_i) = \exp(-\lambda \mathbf{D}_i - \gamma \mathbf{D}_i^2 / n)$ 表示在第 i 个体素单元内的克隆源性细胞在受到 \mathbf{D}_i 剂量的照射后的存活率, λ 为内在放射敏感性参数, 表示不可修复的放射损伤, γ 为两次分次照射之间可以修复的损伤。比值 λ/γ 称为分次剂量校正因子。

Hall 等人^[24]结合 LQ-泊松模型, 同时考虑肿瘤克隆源性细胞后期加速再增殖, 得到更通用的 LQ-泊松模型:

$$TCP(D) = \exp(-N \exp(-D(\lambda + \gamma d) + (t/T_{eff}) \ln 2)) \quad (3.10)$$

式中 d 为单次剂量, D 为总剂量, t 为总治疗时间与增殖滞后时间之差, T_{eff} 为细胞的有效倍增时间。

LQ-泊松模型综合考虑被杀灭肿瘤克隆源性细胞的统计学分布, 分次照射间细胞修复以及细胞后期的加速再增殖, 因而成为目前主流放射治疗系统选用的 TCP 模型之一。遗憾的是, LQ-泊松模型的参数通常是根据离体实验确定的。目前已有足够的临床和实验数据证明, 肿瘤细胞放射敏感性在群体间存在显著的差异, 即使是在同一肿瘤内的不同肿瘤细胞的放射敏感性也存在显著差别, 这些差异将影响剂量效应曲线的斜率。因此, 离体实验得到的模型参数不能直接地应用于临床。在分次照射期间, LQ-模型假设亚致死损伤的细胞获得完全性修复并且没有细胞增殖。目前大多数比值 λ/γ 是来自于动物实验, 而来自人体组织的精确参数很少, 因此在临床应用上应该慎重。比值 λ/γ 还与不同的分次照射方案有关, 尤其在晚反应组织中, 对等效剂量估算的不准确程度将明显上升。肿瘤组织存在多种不确定的内在因素也可以影响 λ/γ 值, 如坏死、乏氧等。

3) Zaider-Minerbo 模型

前面介绍的 TCP 模型主要利用描述临床观察结果的方程式来估算靶区肿瘤控制率。放射生物学家一直在追求以肿瘤细胞放射效应为基础的 TCP 模型。Zaider 等人^[25]提出了一种理论上非常严谨的 TCP 模型, 此模型综合考虑了细胞放射效应的杀灭和增殖两种因素。但是, 由于 Zaider-Minerbo 模型为积分的形式, 不方便应用到常规的分次照射中, Stavreva 等人^[22, 26]改进了 Zaider-Minerbo 模型, 将其离散化, 以方便计算机处理:

$$TCP(t = T_n) = \left[1 - \frac{p_s(T_n) e^{\lambda T_n}}{\left[1 - p_s(T_n) e^{\lambda T_n} \sum_{k=1}^n p_s^{-1}(T_{k-1}) (e^{-\lambda T_k} - e^{-\lambda T_{k-1}}) \right]} \right]^N \quad (3.11)$$

其中 T_k 为第 k 次照射和第 1 次照射的时间间隔, T_n 为总治疗时间, $p_s(T_k)$ 为细胞经第 k 次照射后的存活率。假设亚致死损伤的细胞可以完全修复, 用 LQ 模型预测 $p_s(T_k)$, 则可得:

$$p_s(T_k) = \exp \left[-\lambda \left(\frac{k}{n} D \right) - \gamma \frac{\left(\frac{k}{n} D \right)^2}{k} \right] \quad (3.12)$$

对于非均匀剂量照射:

$$TCP(\mathbf{D}) = \prod_i TCP(\mathbf{D}_i, \mathbf{v}_i) \quad (3.13)$$

式中

$$TCP(\mathbf{D}_i, \mathbf{v}_i) = \left[1 - \frac{p_s^i(T_n) e^{\lambda T_n}}{\left[1 - p_s^i(T_n) e^{\lambda T_n} \sum_{k=1}^n \frac{1}{p_s^i(T_k)} (e^{-\lambda T_k} - e^{-\lambda T_{k-1}}) \right]} \right]^{N \mathbf{v}_i} \quad (3.14)$$

而 $p_s^i(T_k)$ 为:

$$p_s^i(T_k) = \exp \left[-\lambda \left(\frac{k}{n} \mathbf{D}_i \right) - \gamma \frac{\left(\frac{k}{n} \mathbf{D}_i \right)^2}{k} \right] \quad (3.15)$$

Zaider-Minerbo 模型综合考虑细胞放射杀灭效应和增殖的随机过程理论而得到,适用于计算目前所有照射方式上所得剂量分布的 TCP 值。事实上, Zaider-Minerbo 模型概括了之前给出的多种 TCP 模型^[27-30]。针对不同照射方式对模型参数进行适当的调整, Zaider-Minerbo 模型就可以退化为相应照射方式下的 TCP 模型,因而 Zaider-Minerbo 模型最有可能成为治疗系统中的通用 TCP 模型。

3.2.2.3 NTCP 准则

NTCP 模型是用来描述在正常组织出现并发症的概率。大量证据表明,正常组织的放射反应与接收照射的正常组织体积有关。因此,正常组织的受照体积是反映组织并发症的一个特别重要的参数。体积效应的程度依赖于不同组织内部的结构体系。针对不同类型的体积效应,已有学者提出了一些模型^[19, 20, 31-35]。在这些模型中,一些是利用描述临床观察结果的方程式来估算正常组织并发症的经验模型,而另一些是利用器官放射生物学机理进行预测的模型。

1) Schultheiss 模型

Schultheiss 于 1983 年根据广义线性模型推导出正常组织在均匀剂量 D 照射下与 NTCP 关系的逻辑模型^[19]:

$$NTCP(D) = \left[1 + \left(\frac{TD_{50}(1)}{D} \right)^k \right]^{-1} \quad (3.16)$$

其中 $TD_{50}(1)$ 为剂量效应曲线中 $NTCP = 0.5$ 时的剂量值, k 为描述剂量效应曲线特征的一个参数,与剂量效应曲线的斜率有关。 $TD_{50}(1)$ 可以用最大的损伤容许剂量 $TD_{50/5}$ 替代, $TD_{50/5}$ 为标准治疗条件下,放射治疗后 5 年内因放射治疗造成的严重放射损伤的患者不超过 50% 时所使用的照射剂量。考虑到临床中使用非均匀剂量照射正常组织, Niemierko^[20] 提出用 EUD 来替换均匀剂量 D , 以此引入体积效应因子和病人的 DVH 信息。Schultheiss 逻辑模型只有两个参数,形式简单直观,便于计算。引入 EUD 后弥补了无法预测非均匀照射下放射生物效应的缺陷,因此在临床上非常受欢迎。

2) Lyman-Kutcher-Burman(LKB)模型

Lyman^[31]首先提出了 S 形剂量效应(sigmoidal dose response, SDR)积分模型, 用来描述正常组织的全部或部分体积 V 受到均匀剂量 D 照射后的剂量效应:

$$NTCP(D) = \Phi\left(\frac{D - D_{50}(V)}{mD_{50}(V)}\right) \quad (3.17)$$

式中 $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x \exp\left(-\frac{t^2}{2}\right) dt = \frac{1}{2} \left[1 + \operatorname{erf}\left(\frac{x}{\sqrt{2}}\right) \right]$ 为标准正态累积分布函数, $D_{50}(V) = D_{50}(V=1)V^n$,

$D_{50}(V=1)$ 、 $D_{50}(V)$ 分别为全部体积和部分体积 V 受照射时, 由放射损伤引起的 NTCP 值为 50% 时所需的剂量, n 为体积效应因子, m 为控制 NTCP 剂量效应曲线斜率的参数。在临床治疗中, 正常组织接受到非均匀剂量的照射。为了能预测非均匀剂量分布下正常组织的生物效应, Kutcher 和 Burman^[32]利用 gEUD 代替 SDR 模型中的均匀剂量 D , 形成了目前通用的 Lyman-Kutcher-Burman (LKB)模型:

$$NTCP(\mathbf{D}) = \Phi\left(\frac{gEUD(\mathbf{D}) - D_{50}(1)}{mD_{50}(1)}\right) \quad (3.18)$$

LKB 模型为包含三个参数的经验模型, 其形式比 Schuhheiss 模型复杂, 但是其预测结果的精度要高于 Schuhheiss 模型。商用治疗计划系统通常选用 LKB 模型计算 NTCP 值, 如 Pinnacle 计划系统。

3) 临界体积模型

与以上两种利用描述临床观察结果的方程式来估算出现正常组织并发症概率的经验模型相比, 基于器官放射生物学机理的 NTCP 模型可以更好地预测正常组织的放射反应。假设正常器官是由多细胞的 FSU 组成, 整个器官的放射反应取决于每个 FSU 的放射敏感性以及这些 FSU 组成该器官功能的方式^[33], 当大量 FSU 被破坏且超过功能储备分布描述的阈值时, 器官就会表现出并发症。每个 FSU 的敏感性与其组成细胞的放射敏感性及其增殖属性有关。现有三种假设来模拟 FSU 组成该器官功能的方式: 串行反应、并行反应和等级反应。根据不同的 FSU 组成该器官功能的方式, Niemierko 等人^[34]提出了适用于串行器官的临界体积(critical volume, CV)模型, Stavrev 等人^[35]在 CV 模型的基础上进行修改, 使其也能适用于并行器官的模型:

$$NTCP(\mathbf{D}) = \Phi\left(\frac{-\ln(-\ln \bar{\mu}_d) - \ln(-\ln \mu_{cr})}{\sigma}\right) \quad (3.19)$$

式中 $\Phi(x)$ 为标准正态累积分布函数, $\bar{\mu}_d$ 为平均相对损伤体积,

$$\bar{\mu}_d = \sum_i v_i p_{FSU}(\mathbf{D}_i) = \sum_i v_i \Phi\left(\sqrt{2\pi} \gamma_{50}^{FSU} \ln \frac{\mathbf{D}_i}{D_{50}^{FSU}}\right) \quad (3.20)$$

其中 γ_{50}^{FSU} 和 D_{50}^{FSU} 是 FSU 的剂量效应参数, μ_{cr} 为 FSU 的临界分数, σ 表示细胞的放射敏感差异性。

将正常组织假设为由多个 FSU 组成, 还需要考虑以下三点: 1) 同一器官内不同 FSU 的放射敏感性也不同; 2) 对于不同的人, FSU 的放射敏感性和功能储备分布也可能不同; 3) 剂量分布的不确定性。因此, 虽然临界体积模型是基于器官的放射生物学机理的 NTCP 模型, 但是由于临界体积模型的参数非常难确定且对高剂量分布不敏感, 不利于在临床上广泛应用。

3.3 使用生物准则的优点

放射治疗的目的是给予靶组织放射治疗所需剂量，同时最大限度地减少发生正常组织并发症的风险。一直以来物理师或医生使用物理准则来评估放射治疗计划的质量，即剂量分布和 DV 参数。DV 准则只是测量生物效应的一种替代措施。因此，为了在治疗过程中能够更精确地反应是否达到临床治疗目的，应该使用生物准则来替代 DV 准则。

3.3.1 基于剂量-体积准则的治疗计划的缺点

单个或多个 DV 约束能够用于逆向治疗计划或评估计划是基于临床研究证实的治疗结果(如 TCP 或 NTCP) 与特殊 DV 约束之间的关系。例如，在肺组织中， V_{20} (该组织内接收到剂量大于等于 20Gy 的体积占该组织总体积的百分比)用来衡量一个放射治疗计划造成等级为 2 或 3 放射性肺炎的概率。使用 DV 准则评价治疗方法的缺点如下：

1) 通常情况下，DVH 上的多个点与组织并发症有关，如在肺组织中考考虑 V_5 , V_{40} 和平均剂量。这表明 DVH 曲线上的不同部分可能与出现并发症的概率有关。而且，这种相关性还与所使用的放射治疗技术有关，如 IMRT, 3D 适形放射治疗等^[10]。

2) 一般来说，在逆向计划中是间接地使用 DV 约束。为达到最佳的 TCP 和 NTCP 值，需要使用大量的技术和经验来选择恰当的 DV 约束值及其对应权重值。通常情况下，使用 1-3 个 DV 约束条件来控制一个危险器官的 DVH 曲线的分布。但是，满足这些约束条件的 DVH 曲线有很多条，所对应的 NTCP 的值也不同。在基于生物模型的优化过程中，利用生物准则函数取代 DV 约束条件来更加有效地驱动 DVH 曲线的变化，获得最优计划，以实现最有利的整体治疗效果，而不只是满足几个控制点处的约束条件。

3) 在逆向优化过程中指定多个 DV 约束条件，增加了优化算法的复杂度。而且，目标函数不再是凸函数，导致目标函数有多个局部最小值^[11, 12]。这意味着，专门为求解全局最小值的优化问题而设计的搜索算法很可能被困在局部的最小值处，从而存在优化结果是局部最优解的风险。

目前，大多数的临床放射治疗计划系统采用基于 DV 约束条件的优化算法，而且也没有根据 DVH 矩阵计算基于生物学指标的评估工具。在开发与输出结果相关的公共数据集和比较不同的放射治疗方法中，利用生物学指标进行计划评估是很重要的。在逆向计划中，采用 DV 约束驱动所得剂量分布和采用生物约束条件驱动所得剂量分布的差别很大，优先采用以生物学为基础的优化评估工具是非常有必要的。

3.3.2 基于生物准则的治疗计划的优点

与 DV 准则相比，基于生物模型的优化准则更灵活、更直接地与实际治疗效果相关。构造基于生物效应的模型来估算肿瘤或正常组织对剂量分布的响应，不仅需要在临床数据范围内做一些推断，而且还需要在临床证据范围外进行一些额外的推断。模型中所使用的参数是通过临床采集的基本数据集进行处理得到的。但是，如果实际剂量分布大大偏离获取生物模型参数所使用的基本数据集，不能保证基于生物效应的模型一定能够正确地估算肿瘤或正常组织对该剂量分布的响应。然而，在方案优化中只要使用生物模型就可以将优化引向更有利的剂量分布，可以暂且不考虑生物模型所使用参数是否有效^[13]。

基于生物准则的方案优化算法的另一个优点是，能够克服由于不同患者之间不可避免的解剖结构差异以及相应剂量分布的变化所引起的方案优化质量变差。在这方面，由多个 DV 准则控制一个器官的 DVH 曲线的逆向优化算法可能会迅速出错，因为需要给出这些 DV 准则的优先级而且 DV 约束条件的处方量也应该随患者解剖结构的改变而变化。例如，对于膀胱壁， $V_{65}<50\%$ 是其中的一个 DV 控制点^[14]，但是在一些患者中膀胱壁的 V_{65} 可以明显地低于 50%，因此在这些患者中使用 $V_{65}<50\%$ 为约束条件显然不能达到最优解。而且这些 DV 准则最好能被一个表示该器官对当前剂量分布反应的品质数值表示，以避免和其它器官混淆。相比之下，生物模型有将同一个器官的多个 DV 准则按固有的优先次序结合成一个品质数值的

潜力。

3.4 等效凸准则

直接将(非凸的)准则作为目标函数和约束函数会导致优化问题变为非凸优化问题,增加了求解优化问题的难度。求解优化问题的算法可以被分为两类:基于梯度的方法和随机方法。基于梯度的方法可以在很短的时间内找到最优解。但是基于梯度的方法要求目标函数是凸函数,这限制了构成目标函数的优化准则的选取。随机搜索方法不要求目标函数为凸函数,如模拟退火方法。因此,在随机方法中目标函数所使用的优化准则不限于凸准则。但是,使用随机方法寻找全局最优点的速度比较慢。本章中,我们使用基于梯度的方法解优化问题以缩短方案优化的时间。然而,部分物理准则和生物准则为非凸函数,限制了直接利用这些准则构造目标函数。已有一些学者提出了不同的方法将非凸准则转换为等效凸准则^[5, 8, 36, 39, 40]。这一节中,我们主要介绍这些已有的转换方法。

3.4.1 等效凸物理准则

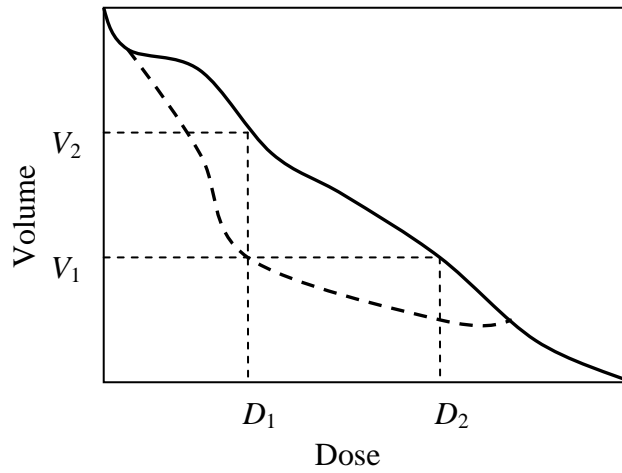


图 3-1 DV 约束函数的 DVH 图示

按实际剂量与处方剂量或耐受剂量的二次偏差形式给出的最小剂量准则,最大剂量准则和平均剂量准则都是凸准则,可以直接作为子目标函数。然而由于 DVH 准则不能直接表示为二次偏差的形式,因此需要将其转换为等效凸准则。DVH 约束条件定义为 $V(>D_1) < V_1$,即器官内接受到剂量大于 D_1 的体积小于 V_1 。如图 3-1 所示,正常组织接受到的实际剂量(实线)比理想剂量分布(虚线)高,而我们希望实际剂量曲线至少能通过 DVH 控制点 (D_1, V_1) 。因此,由实际剂量曲线和两条边界(从点 (D_1, V_1) 到点 (D_2, V_1) 以及从点 (D_1, V_1) 到点 (D_1, V_2)) 所组成多边形内的剂量分布不满足 DVH 约束条件。Wu 等人^[36]提出了等效准则的方法来惩罚不满足约束条件的剂量分布,且保证准则为凸函数:

$$f_{DVH}(\mathbf{D}) = \frac{1}{N} \sum_{i=1}^N H(\mathbf{D}_i - D_1) \cdot H(D_2 - \mathbf{D}_i) \cdot (\mathbf{D}_i - D_1)^2 \quad (3.21)$$

随后有学者^[37, 38]提出,利用这种方法得到的等效 DVH 准则是非凸函数,而且将等效 DVH 准则应用到方案优化中会使优化问题在局部收敛。但是他们^[37, 38]同时又指出,这只是在理论上存在的一个问题,似乎对实际临床应用没有影响。

3.4.2 等效凸生物准则

Wu 等人^[39]用组合函数的形式分别给出了靶目标和正常组织的等效凸 EUD 函数。Choi 等人^[40]证明了,当 $a \geq 1$ 时, gEUD 是凸函数,可以直接用来构造目标函数,当 $a < 1$ 时, gEUD 是凹函数,求反之后即可用来构造目标函数。Romeijn 等人^[5]证明了,如果非凸准则可以表示

为一个单调递增函数和一个凸准则的复合形式，即

$$F_k(x) = (h_k \circ G_k)(x) \equiv h_k(G_k(x)) \quad (3.22)$$

式中 F_k 为非凸生物准则， h_k 为单调递增函数， G_k 为凸准则，那么可以用该凸准则取代非凸准则，且可以在多目标优化问题中产生一个等效的 Pareto 有效边界。表 3.1 为使用该方法给出的部分非凸生物准则的分解结果。

表 3.1 部分非凸生物准则的分解结果

准则	参数范围	分解结果	
		G_k	$h(z)$
EUD(式 3.4)	$a > 0$	-EUD	z
		F	$(1/a)\ln(z)$
gEUD(式 3.5)	$-\infty < a \leq 0$	-gEUD	z
	$-\infty < a < 0$	F^g	$-z^{1/a}$
	$a = 0$	F^g	$-e^{-z}$
	$1 \leq a < \infty$	gEUD	z
	$1 \leq a < \infty$	F^g	$z^{1/a}$
TCP(式 3.8)	$N, a > 0; n \geq 1; \lambda, \Delta T \geq 0$	$-\ln(\text{TCP})$	$-e^{-z}$
		-EUD	$-\exp(-Ne^{\lambda(n-1)\Delta T - az})$
NTCP(式 3.18)	$1 \leq a \leq \infty; m, D_{50} > 0$	gEUD	$\Phi((z - D_{50})/mD_{50})$
NTCP ^[41]	$1 \leq a < \infty, \Delta > 0$	gEUD	$1 - \exp((-z/\Delta)^a)$

*注: $F(\mathbf{D}) = \frac{1}{N} \sum_{j=1}^N f(\mathbf{D}_j)$, $f(\mathbf{D}_j) = \exp(-a\mathbf{D}_j)$; 若 $f(\mathbf{D}_j)$ 为表示为分段函数: $f^g(\mathbf{D}_j) = \begin{cases} \mathbf{D}_j^a & a \neq 0 \\ -\ln \mathbf{D}_j & a = 0 \end{cases}$, 则

函数 F 变为 F^g 。

Hoffmann 等人^[8]已经证明，对于生物准则，利用严格递增函数可以将任意一个求解最小优化问题中的非凸准则转换为等效凸准则，并且导致多目标优化问题产生一个等效的 Pareto 有效边界，其转换公式为：

$$G_k(x) = (h_k \circ F_k)(x) \equiv h_k(F_k(x)), \quad (3.23)$$

式中 h_k 为严格递增函数， F_k 为非凸生物准则， G_k 为等效凸生物准则。还需注意，如果 h_k 是凸函数且单调递增(单调递减)， F_k 是凸函数(凹函数)，那么 G_k 为凸函数；如果 h_k 是凹函数且单调递增(单调递减)， F_k 是凹函数(凸函数)，那么 G_k 为凹函数。表 3.2 为使用该方法给出的部分基于 LQ-泊松模型的非凸生物准则的分解结果。

表 3.2 部分基于 LQ-泊松模型的非凸生物准则的变换公式

准则	参数限制	$h(z)$
gEUBED ^[8]	$a \geq 1$	z
TCP _{LQ} (式 3.9)	$\mathbf{D}_i > \sqrt{\frac{1}{2} \frac{(\alpha/\beta)n}{\alpha}} - \frac{1}{2}(\alpha/\beta)n$	$-\ln(z)$
TCP _L ^[8]	无	$-\ln(z)$ 或 $\ln(-\ln(z))$
TCP _{pop} ^[42]	$\mathbf{D}_i > \sqrt{\left[\frac{1}{2}(\alpha/\beta)n\right]^2 + \frac{1}{2} \frac{(\alpha/\beta)n}{\alpha}} - \frac{1}{2}(\alpha/\beta)n$	$-\ln(z)$
NTCP _{LQ} (式 3.18)	无	$-\ln(1-z)$

NTCP _{RS} ^[43]	$0 < s < 1$	$-\ln(1-z)$ 或 $-\ln(1-z^s)$
NTCP _{gEUBED} ^[8]	无	$-\ln(1-z)$

表 3.1 和表 3.2 给出了部分生物准则的等效凸准则。从表中可看出，部分生物准则是先取负数，然后在再将其进行分解或转换。因为，这些生物准则本身是凹函数或者等效准则是凹函数，这时在对这些凹函数取负数，就变成了凸函数。将生物准则变换为等效的凸函数有不同的分解或变换方法且生成的等效凸函数也不一定相同。如表 3.2 所示，部分生物准则给出了两个变换公式。同一生物准则有不同的等效凸函数，且其函数特性也有差别。一些等效凸函数的“凸”的程度要比另一些等效凸函数小一些。这个性质对使用分段线性上边界和下边界的方法求 Pareto 等效边界尤为重要，利用“凸”的程度较小的凸函数可以得到更紧致的 Pareto 等效边界^[44]。

3.5 目标函数

在本章中，我们只考虑 FMO 问题。在优化过程中，剂量分布可以表示为 $\mathbf{D}=\mathbf{W}\mathbf{x}$ ：

$$\begin{bmatrix} \mathbf{D}_1 \\ \mathbf{D}_2 \\ \vdots \\ \mathbf{D}_N \end{bmatrix} = \begin{bmatrix} \mathbf{W}_{1,1} & \mathbf{W}_{1,2} & \cdots & \mathbf{W}_{1,m_1} & \cdots & \cdots & \mathbf{W}_{1,(m_{k-1}+1)} & \mathbf{W}_{1,(m_{k-1}+2)} & \cdots & \mathbf{W}_{1,m_k} \\ \mathbf{W}_{2,1} & \mathbf{W}_{2,2} & \cdots & \mathbf{W}_{2,m_1} & \cdots & \cdots & \mathbf{W}_{2,(m_{k-1}+1)} & \mathbf{W}_{2,(m_{k-1}+2)} & \cdots & \mathbf{W}_{2,m_k} \\ \vdots & \vdots & & \vdots & & & \vdots & \vdots & & \vdots \\ \vdots & \vdots & & \vdots & & & \vdots & \vdots & & \vdots \\ \mathbf{W}_{N,1} & \mathbf{W}_{N,2} & \cdots & \mathbf{W}_{N,m_1} & \cdots & \cdots & \mathbf{W}_{N,(m_{k-1}+1)} & \mathbf{W}_{N,(m_{k-1}+2)} & \cdots & \mathbf{W}_{N,m_k} \end{bmatrix} \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \\ \vdots \\ \mathbf{x}_{m_1} \\ \vdots \\ \vdots \\ \mathbf{x}_{(m_{k-1}+1)} \\ \vdots \\ \vdots \\ \mathbf{x}_{m_k} \end{bmatrix} \quad (3.24)$$

\mathbf{W} 为剂量效应矩阵，其含义为单位强度射束在待照射体模内的剂量分布数据； \mathbf{x} 为注量矩阵，是所有照射野强度分布矩阵，即各照射野方向的注量图分布，构成的列向量。公式 3.24 描述了由 k 个射野对体模进行照射，得到体模内 N 个体素处的剂量分布矩阵。其中，注量矩阵 \mathbf{x} 包含了 k 个射野的注量图分布， $m_{l-1}+1$ 到 m_l 表示第 l 个射野内的 $m_l - m_{l-1}$ 个射束的强度分布。因此，剂量效应矩阵 \mathbf{W} 和注量矩阵 \mathbf{x} 的乘积就表示在每个射野方向使用不同强度注量分布照射体模得到的整个体模空间内的剂量分布。在 FMO 问题中，通过控制每个射束的强度，射线的能量集中在靶区且尽可能地避开靶区周围的正常组织，从而使得靶区总照射剂量在足以杀灭肿瘤细胞、控制肿瘤扩散的同时，尽量降低周围正常组织的受照射剂量，从而降低并发症发生的概率。因此，FMO 问题的最终目的是得到恰当的注量矩阵 \mathbf{x} 。而解决 FMO 问题是通过建立关于注量矩阵 \mathbf{x} 的目标函数，利用优化算法控制靶区及周围正常组织的剂量分布，从而找到最优的注量矩阵 \mathbf{x} 。

利用前面提到的优化准则便可构造出总的目标函数，而目标函数有不同的构造方法。例如，将部分准则构成目标函数，其余的准则组成限制条件^[2]，这样构造出来的目标函数只能使用规划类的方法解优化问题。在本章中，我们使用梯度类的方法解优化问题，因此只构造一个目标函数，而不需要额外的限制条件。最常见的方法是将一些准则构成子目标函数，然后相加^[45]或相乘^[39, 46, 47]这些加权的子目标函数构成总的目标函数，在这些子目标函数前乘以权重来表示对应的子目标函数的重要性。本章中采用由子目标函数加权的形式构成总目标函数：

$$f(\mathbf{D}(\mathbf{x})) = \sum_{l=1}^L \xi_l f_l(\mathbf{D}(\mathbf{x})) \quad (3.25)$$

式中 f_l 为子目标函数， ξ_l 为权重系数， L 为所有子目标函数的个数。对同一个器官，可以使

用一个或多个子目标函数来控制这个器官内的剂量分布。在 FMO 问题中，剂量分布是关于剂量效应矩阵 \mathbf{W} 和注量图强度矩阵 \mathbf{x} 的线性函数，且在优化过程中 \mathbf{W} 是不可变的。因此，前面介绍的关于剂量分布为凸函数的准则，也同样是关于注量图强度矩阵 \mathbf{x} 的凸准则。通常，子目标函数是由准则和处方量的差构成，并乘以该差值的阶跃函数，表示只惩罚不满足该准则的注量分布。本章中介绍的等效凸物理准则是由实际剂量分布和处方剂量的差乘以该差值的阶跃函数构成，因此可以直接作为子目标函数使用。而等效生物准则没有给出与处方量的关系，因此不能直接使用，需要重新构造。由于在本章中使用基于梯度的方法，L-BFGS 算法^[9]，求解优化问题，因此还需给出子目标函数关于注量图强度矩阵 \mathbf{x} 的一阶导数。一下给出前面介绍的等效物理准则的一阶导数和等效凸生物准则的子目标函数极其一阶导数。

3.5.1 物理子目标函数

由前面介绍可知，最小剂量子目标函数、最大剂量子目标函数和平均剂量子目标函数可以直接由其对应的物理准则构成，无需改变。DVH 子目标函数可以直接由前面介绍的等效凸 DVH 准则构成。下面依次给出这四种子目标函数关于注量图强度矩阵 \mathbf{x} 的一阶导数：

$$\frac{\partial f_{\min}}{\partial \mathbf{x}_j} = \frac{2}{N} \sum_{i=1}^N H(D_{\min} - \mathbf{D}_i) \cdot (\mathbf{D}_i - D_{\min}) \mathbf{W}_{i,j} \quad (3.26)$$

$$\frac{\partial f_{\max}}{\partial \mathbf{x}_j} = \frac{2}{N} \sum_{i=1}^N H(\mathbf{D}_i - D_{\max}) \cdot (\mathbf{D}_i - D_{\max}) \mathbf{W}_{i,j} \quad (3.27)$$

$$\frac{\partial f_{mean}}{\partial \mathbf{x}_j} = \frac{2}{N} \sum_{i=1}^N (\mathbf{D}_i - D_{mean}) \mathbf{W}_{i,j} \quad (3.28)$$

$$\frac{\partial f_{DVH}}{\partial \mathbf{x}_j} = \frac{2}{N} \sum_{i=1}^N H(\mathbf{D}_i - D_1) \cdot H(D_2 - \mathbf{D}_i) \cdot (\mathbf{D}_i - D_1) \mathbf{W}_{i,j} \quad (3.29)$$

3.5.2 生物子目标函数

1) EUD 子目标函数

我们利用由 Choi 等人^[40]提出的等效凸 gEUD 准则构造 EUD 子目标函数，即当 $a \geq 1$ 时，gEUD 是凸函数，当 $a < 1$ 时，gEUD 是凹函数，需求反之后为凸函数。为了控制器官(靶目标和正常组织)内的高剂量分布，通常使用较大的 a 值($a \geq 1$)来反映器官内的高剂量分布，子目标函数用来惩罚高于处方量 $gEUD_0$ 剂量分布：

$$f_{gEUD}(\mathbf{D}) = H(gEUD(\mathbf{D}) - gEUD_0) \cdot (gEUD(\mathbf{D}) - gEUD_0) \quad (3.30)$$

其一阶导数为：

$$\frac{\partial f_{gEUD}}{\partial \mathbf{x}_j} = H(gEUD(\mathbf{D}) - gEUD_0) \cdot \frac{gEUD(\mathbf{D})}{\sum_{i=1}^N \mathbf{D}_i^a} \sum_{i=1}^N \mathbf{D}_i^{a-1} \mathbf{W}_{i,j} \quad (3.31)$$

为了控制器官(靶目标)内的低剂量分布，通常使用较小的 a 值($a < 1$)来反映器官内的低剂量分布，这时子目标函数用来惩罚低于处方量 $gEUD_0$ 剂量分布：

$$f_{gEUD}(\mathbf{D}) = H(gEUD_0 - gEUD(\mathbf{D})) \cdot (gEUD_0 - gEUD(\mathbf{D})) \quad (3.32)$$

其一阶导数为：

$$\frac{\partial f_{gEUD}}{\partial \mathbf{x}_j} = -H(gEUD_0 - gEUD(\mathbf{D})) \frac{gEUD(\mathbf{D})}{\sum_{i=1}^N \mathbf{D}_i^a} \sum_{i=1}^N \mathbf{D}_i^{a-1} \mathbf{W}_{i,j} \quad (3.33)$$

使用不同的 a 值, EUD 子目标函数可以控制器官内的高剂量分布($a \geq 1$)或低剂量分布($a < 1$)。上面给出了两个 EUD 子目标函数, 在实际应用过程中需要根据 a 值的大小选择恰当的 EUD 子目标函数。

2) TCP 子目标函数

Hoffmann 等人^[8]提出将基于 LQ-泊松模型的 TCP 准则(式 3.9)通过严格递增函数 $-\ln(z)$ 变换为等效凸准则。在本章中, 我们使用该等效凸准则构造 TCP 子目标函数, 用来惩罚靶区肿瘤控制率低于 TCP_0 的剂量分布:

$$f_{TCP}(\mathbf{D}) = H(TCP_0 - \ln(TCP(\mathbf{D}))) \cdot (TCP_0 - \ln(TCP(\mathbf{D}))) \quad (3.34)$$

其一阶导数为:

$$\frac{\partial f_{TCP}}{\partial \mathbf{x}_j} = -\frac{N_0}{N} H(TCP_0 - TCP(\mathbf{D})) \sum_{i=1}^N SF_{LQ}(\mathbf{D}_i)(\alpha + 2\beta \mathbf{D}_i / n) \mathbf{W}_{i,j} \quad (3.35)$$

3) NTCP 子目标函数

同样, 利用 Hoffmann 等人^[8]提出的变换方法, 将基于 LKB 模型的 NTCP 准则(式 3.18)通过严格递增函数 $-\ln(1-z)$ 变换为等效凸准则。利用该等效凸准则构造 NTCP 子目标函数来惩罚正常组织内过高的剂量分布, 将该组织的并发症概率控制在 $NTCP_0$ 内:

$$f_{NTCP}(\mathbf{D}) = H(\ln(1 - NTCP_0) - \ln(1 - NTCP(\mathbf{D}))) \cdot (\ln(1 - NTCP_0) - \ln(1 - NTCP(\mathbf{D}))) \quad (3.36)$$

其一阶导数为:

$$\frac{\partial f_{NTCP}}{\partial \mathbf{x}_j} = H\left(\frac{\ln(1 - NTCP_0)}{\ln(1 - NTCP(\mathbf{D}))}\right) \frac{1}{\sqrt{2\pi m D_{50}} (1 - NTCP(\mathbf{D}))} e^{-t^2} \frac{gEUD(\mathbf{D})}{\sum_{i=1}^N \mathbf{D}_i^a} \sum_{i=1}^N \mathbf{D}_i^{a-1} \mathbf{W}_{i,j} \quad (3.37)$$

式中 $t = \frac{gEUD(\mathbf{D}) - D_{50}}{\sqrt{2m D_{50}}}$ 。

3.6 基于等效凸 NTCP 准则的 IMRT 方案优化算法研究

在 FMO 问题中, 已有学者利用(g)EUD 准则构造子目标函数来控制靶区或正常组织的剂量分布^[39, 40, 48-50]。在已知(g)EUD 准则中的参数 a 的情况下, (g)EUD 模型具有为治疗计划提供可靠评价的潜力。但是, (g)EUD 模型不能直接给出正常组织出现并发症的概率。而 NTCP 准则可以直接计算当前剂量分布下正常组织出现并发症的概率, 因此, 相比(g)EUD 模型, NTCP 模型更易于临床应用。在基于 NTCP 准则的逆向计划中, 如果目标函数中所有条件都满足了, 则此时的剂量分布可被接受, 不需要再寻找更优的解。而在基于物理准则或基于(g)EUD 准则的方案优化中, 由于这些准则不能直接估算正常组织出现并发症的概率, 因此, 在优化算法条件都满足的情况下还有必要寻找更优的解。例如, 我们使用一个 NTCP 准则控制一个正常组织的剂量分布, 如果当前剂量分布的 NTCP 值小于处方量, 当前剂量分布可接收就不用寻找更优解; 然而使用多个 DVH 准则控制一个正常组织的剂量分布, 虽然当前剂量分布满足多个 DVH 约束条件, 但我们不知道当前剂量分布的 NTCP 值, 无法确定当前解是否合适, 只能继续寻找更优的解。但是在使用 NTCP 模型之前需要花费大量精力来校正模型中所用参数的值。

一些学者已将 NTCP 准则应用到方案优化中^[2, 45, 46]，而且在一些计划系统中也使用了基于 NTCP 准则的方案优化算法^[13, 45]，如 Varian Eclipse 系统(V. 10)。Mohan 等人使用 TCP 准则和 NTCP 准则构造目标函数，分别利用 TCP 准则和 NTCP 准则控制靶区和正常组织的剂量分布，并使用模拟退火算法解 FMO 问题^[46]。Wang 等人使用相同的目标函数，但使用最速下降法解优化问题^[47]。他们没有讨论目标函数的非凸问题，因此直接使用该方法进行优化是值得商榷的。Stavrev 等人在逆向优化问题中使用基于物理-生物准则的混合模型，并使用非线性规划算法解优化问题^[2]。在混合模型中，目标函数为一些 NTCP 准则的加权和，用来控制正常组织内的剂量分布；在约束条件里利用最大剂量和最小剂量对靶区剂量进行控制。如前文所述，NTCP 准则是关于剂量分布的 S 形函数，因此，该混合模型是关于注量图强度矩阵 \mathbf{x} 的非凸函数。在方案优化中直接使用非凸的生物准则会增加求解优化问题的难度。为了克服这一缺点，Hoffmann 等人和 Romeijn 等人分别提出了不同的将生物准则转换为等效凸准则的方法，并讨论了满足凸函数所需的条件。但是，他们只是在理论上证明了这种方法的可行性，没有在实际中应用。在本章中，我们使用等效凸 NTCP 准则构造目标函数，并在前列腺肿瘤病人数据上验证提出算法的有效性。

3.6.1 构造总目标函数

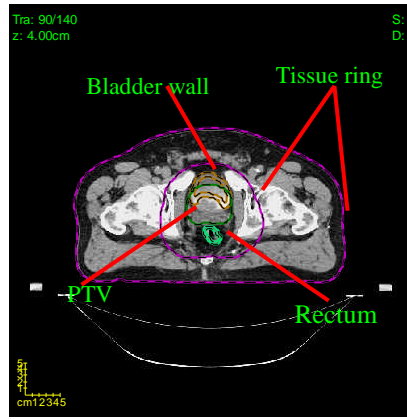


图3-2 前列腺患者内各器官的结构分布。

TCP 准则主要用来反映靶区的剂量不足问题。当靶区出现热点或整体的剂量偏高时，使用基于 TCP 模型的子目标函数不能及时地将高剂量信息反馈给目标函数。因此，为了控制靶区剂量分布的均一性，我们使用物理准则来控制靶区剂量分布。我们提出一种基于物理模型-生物模型的混合目标函数：利用一个最小剂量子目标函数 $f_{\min}(\mathbf{D}^{PTV})$ 和一个均匀剂量子目标函数 $f_{\text{mean}}(\mathbf{D}^{PTV})$ 控制靶区剂量分布；每个正常组织内的剂量分布都由一个 NTCP 子目标函数控制，在本章中正常组织只考虑膀胱 $f_{NTCP}(\mathbf{D}^{Bladder})$ 和直肠 $f_{NTCP}(\mathbf{D}^{Rectum})$ ；为了能使 PTV 外区域的高剂量快速下降，减少正常组织内的剂量分布，使用一个 DVH 子目标函数 $f_{DVH}(\mathbf{D}^{Normal})$ 控制剩余正常组织的剂量分布。前列腺肿瘤患者内各器官的解剖结构分布如图 3-2 所示，其中“Tissue ring”表示其余的正常组织。因此，采用前文介绍的构造目标函数的方法(式 3.25)构造前列腺肿瘤病例的总目标函数：

$$f(\mathbf{D}) = \xi_1 f_{NTCP}(\mathbf{D}^{Bladder}) + \xi_2 f_{NTCP}(\mathbf{D}^{Rectum}) + \xi_3 f_{\min}(\mathbf{D}^{PTV}) + \xi_4 f_{\text{mean}}(\mathbf{D}^{PTV}) + \xi_5 f_{DVH}(\mathbf{D}^{Normal}) \quad (3.38)$$

式中 ξ_i 为权重系数。

3.6.2 求解 FMO 问题

本章中，利用基于梯度的优化算法，L-BFGS 算法^[9]，求解约束最小化优化问题，其流程如图 3-3 所示。这类方法是在循环迭代的过程中根据目标函数的梯度信息优化注量图强度矩阵 \mathbf{x} 。在每次迭代过程中，首先利用常规线性搜索方法计算出在当前搜索方向的最大可接受

步长；然后根据前面计算所得步长和当前的 **Hessian** 矩阵计算注量图强度矩阵 \mathbf{x} ，并利用式 3.24 重新计算剂量分布；最后，如果当前剂量分布不满足迭代的终止条件，利用 **L-BFGS** 算法更新 **Hessian** 矩阵，开始进入下一次迭代；如果当前剂量分布满足迭代的终止条件，终止迭代，输出注量图强度矩阵 \mathbf{x} 和剂量分布。其中利用 **Wolfe** 条件法进行常规线性搜索。**Wolfe** 条件法是由一组不等式构成，用来执行非精确线性搜索，特别是在拟牛顿算法中。**L-BFGS** 是拟牛顿类优化算法中的一种。相对于 **BFGS**(Broyden-Fletcher-Goldfarb-Shanno)算法，**L-BFGS** 算法利用近似估算的逆 **Hessian** 矩阵引导搜索的方向，而没有直接计算大规模的 **Hessian** 矩阵，从而减少了优化算法所需内存。正因为 **L-BFGS** 算法在优化过程中使用较少的内存，这种算法在解大规模的优化问题中非常受欢迎。迭代的终止条件有两种：1) 总目标函数的值足够小 ($<1e-6$)；2) 线性搜索方法沿当前的搜索方向找不到更合适的点。在本章的方案优化算法中计算步长和更新 **Hessian** 矩阵都需要计算目标函数的一阶导数。

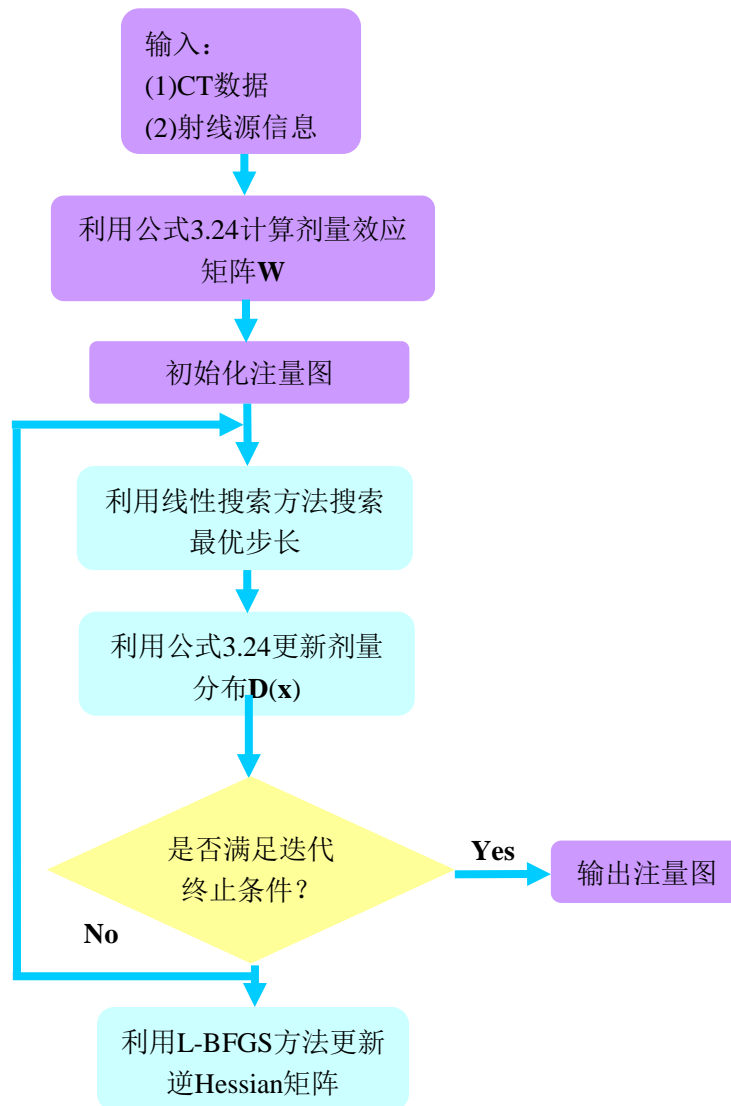


图3-3 本文所用优化算法的流程图。

3.6.3 对比实验

本章提出的方法与其它两种基于NTCP准则的方案优化方法^[2, 46]进行了比较。在Mohan等人^[46]设计的方案优化方法中，目标函数是由TCP准则和NTCP准则的乘积构成。在本章的对比实验中，该优化方法的目标函数由两个NTCP准则和一个TCP准则的乘积构成，分别控制膀胱、

直肠和PTV内的剂量分布。而在Stavrev等人^[2]提出的基于物理-生物准则的混合方案优化模型中,目标函数由两个NTCP子目标函数的和构成,分别反馈膀胱和直肠内的剂量分布信息。靶目标中的剂量分布则在约束条件中由最大剂量和最小剂量进行控制。这两种优化方法直接利用基于LKB模型的NTCP准则构成目标函数,因此其对应的优化问题应使用随机类的优化算法求解。Mohan等人^[46]的方法使用模拟退火方法求解FMO问题,而Stavrev等人^[2]的方法使用非线性规划方法进行求解。在对比实验中,这两种算法所用到TCP模型和NTCP模型中的参数分别来自于这两篇文献,即文献[46]和文献[2]。为了便于区分,这两种方法的计算结果根据其所使用的优化算法分别被标记为“SA”(simulated annealing)和“NP”(nonlinear programming)。本章提出的方法,由于使用了等效凸NTCP准则而被标记为“EC-NTCP”(equivalent convex NTCP criteria)。

本章提出的方法还与基于DVH准则的方案优化方法进行了比较。基于DVH准则的方案优化方法的总目标函数由多个前文所述的DVH子目标函数的和构成,其优化结果被标记为“DVH”。由于其总目标函数是凸函数,因此可以直接使用前面介绍的L-BFGS优化算法求解优化问题。而对于另外两种优化算法,只需将算法流程图(图3-3)中的循环部分分别改为模拟退火算法和非线性规划算法即可。

3.6.4 实验材料及评价准则

本章提出的方案优化方法在10个前列腺肿瘤病例上做了对比实验。患者在仰卧位进行了模拟和治疗。在CT数据上勾画靶区和危及器官(膀胱和直肠),如图3-2所示。直肠壁和膀胱壁是通过将手动圈定的直肠轮廓和膀胱轮廓分别外扩5mm得到的。靶区主要是前列腺,没有考虑盆腔淋巴结。PTV是在CTV的基础上在向后的方向上外扩5mm,在其余的方向上外扩10mm得到。而表示其余正常组织的“Tissue ring”则是将PTV外扩5cm之外的所有组织。我们使用5个6MeV共弧面照射的光子射野照射靶目标,其机架角度分别为:36°,100°,180°,260°和324°。利用CERR开源软件^[51]上的经典笔形束算法^[52]计算剂量效应矩阵 \mathbf{W} 。靶区的处方剂量为78Gy,膀胱和直肠出现并发症的处方量都设置为零,即尽可能地降低正常组织接收到的剂量。膀胱壁所使用的NTCP模型的放射生物参数来自于Dale等人^[53]的文献,而直肠壁所使用的NTCP模型的放射生物参数来自于Peeters等人^[54]的文献,如表3.3中所示。

表 3.3 本章提出方案算法中,不同危及器官所使用 NTCP 模型的放射生物参数。

器官	$n(1/a)$	m	$D_{50}(\text{Gy})$	备注
膀胱	0.13	0.11	62	Severe proctitis/necrosis/stenosis/fistula
直肠	0.13	0.14	80.7	RTOG grade ≥ 3 bladder complications

表 3.4 危及器官的 DV 约束条件

器官	剂量-体积参数			
膀胱	$V_{65} < 50\%$	$V_{70} < 35\%$	$V_{75} < 25\%$	$V_{80} < 15\%$
直肠	$V_{50} < 50\%$	$V_{60} < 35\%$	$V_{65} < 25\%$	$V_{70} < 20\%$ $V_{75} < 15\%$

本章提出的算法及对比实验中所使用的其它方案优化算法都是在MATLAB平台上编程实现的。通过利用文献[55]中介绍的高斯随机数生成函数取代MATLAB代码(simulannealbnd)中默认的随机数生成函数来实现模拟退火算法。而MATLAB代码(fmincon)直接被用来实现非线性规划算法。为了评价不同的方案优化方法,我们利用Marks等人^[56]给出的临床实践指导来评价方案优化的结果,如表3.4所示。同时该指导也被作为基于DVH准则的方案优化方法中DVH子目标函数的参数。我们还使用gEUD(公式3.5)、NTCP(公式3.18)、TCP、迭代次数、计算时间以及最终总目标函数的值对不同方案优化算法的结果进行了评价。非参数检验方法(Wilcoxon test)被用来检测不同方案优化方法之间的相关性。

3.6.5 实验结果

3.6.5.1 NTCP 子目标函数的权重及参数 a 对新算法优化结果的影响

在一例前列腺肿瘤患者上测试了不同的 NTCP 子目标函数权重和不同的参数 a 对新算法优化结果的影响。总目标函数(公式 3.38)的权重依次取 60, 90, 30, 30 和 5; 最小剂量子目标函数的参数 D_{\min} 为 74Gy; 平均剂量子目标函数的参数 D_{mean} 为 78Gy; DVH 子目标函数的约束条件为 $V_{35} < 8\%$; 膀胱和直肠所使用的 NTCP 子目标函数中参数如表 3.3 所示。保持公式 3.38 中其余参数不变而只改变直肠所使用 NTCP 子目标函数的权重。依次取权重 40, 60, 90, 120 和 140, 其优化结果如图 3-4 所示。其中, 蓝色实线是基于 DVH 准则的方案优化方法所得结果, 不同颜色的虚线表示使用不同权重所得的优化结果, 并分别用 weight=40, 60, 90, 120 和 140 进行标注。我们可以从图 3-4 观察到, 直肠壁上的剂量分布随权重系数的增加而逐渐减小。增加控制直肠剂量分布的 NTCP 子目标函数的权重将导致该子目标函数对总目标函数的贡献增加。因此, 总目标函数给与这个危及器官的剂量分布更多的惩罚, 从而减少了该危及器官上的剂量分布。但是, 随着权重的增加, PTV 在总目标函数中的重要性也相对减少了, 因此同时增加了靶区高剂量和低剂量的分布, 减小了靶区剂量分布的均一性。

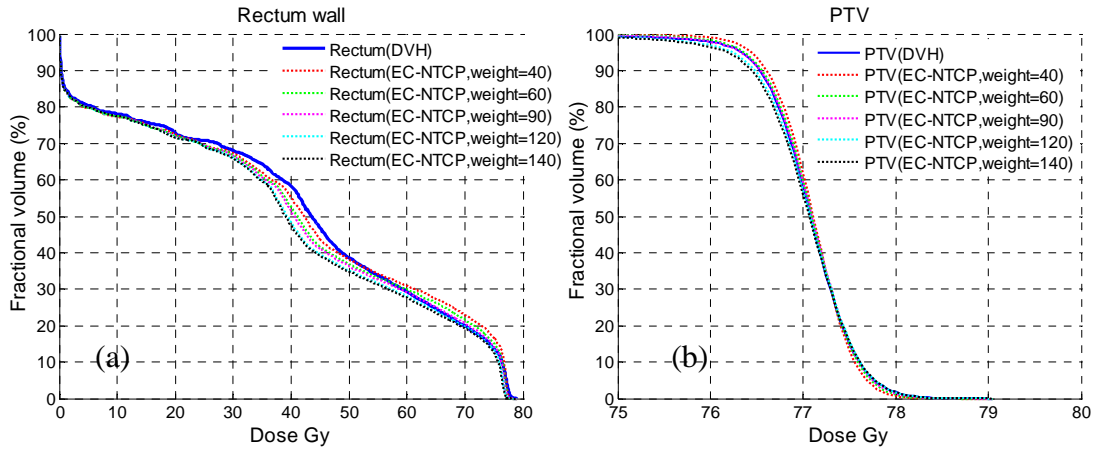


图3-4 在一例前列腺患者上使用不同的权重系数所得优化结果的DVH曲线: (a) 直肠壁; (b) PTV。

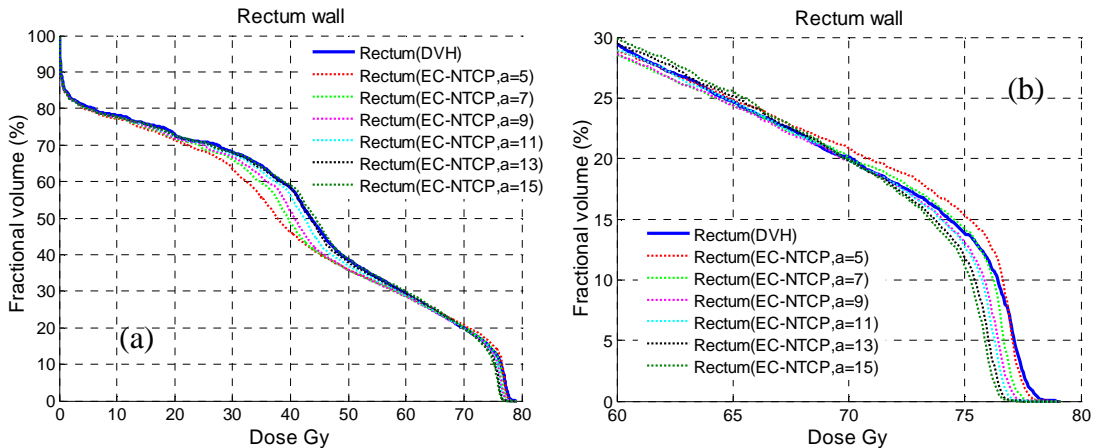


图3-5 在一例前列腺患者上使用不同的参数 a 所得优化结果的DVH曲线: (a) 直肠壁; (b) 直肠壁DVH曲线的高剂量部分。

参数 a 的值可以控制正常组织内主要剂量减少的区域。随着参数 a 值的增加, 低剂量区域的贡献逐渐减少, 高剂量区域被赋予更多的惩罚。保持公式 3.38 中其余参数不变, 只改变所使用的直肠 NTCP 子目标函数的参数, 依次取参数 a 的值为 5, 7, 9, 11, 13 和 15, 其优化结果如图 3-5 所示。其中, 不同颜色的虚线表示在新方法中使用不同的参数 a 值所得优化结果, 并分别用 $a=5, 7, 9, 11, 13$ 和 15 进行标注。从图 3-5 中我们可以观察到, 随着参数

a 值的增加, 危及器官(直肠壁)的主要剂量减少区域从低剂量区域向高剂量区域移动。直肠壁的低剂量区域随着参数 a 值的增加而增加, 如图 3-5(a)所示; 而在高剂量区域内, 随着参数 a 值的增加, 其高剂量分布有明显且稳定的减少, 如图 3-5(b)所示。

3.6.5.2 与基于 DVH 准则的方案优化方法进行比较

本章提出的算法首先在一例前列腺肿瘤患者上与基于 DVH 准则的方案优化方法进行比较, 并使用 DVH 曲线和 NTCP 值对两种方法的优化结果进行评估。因为这一例患者与前面讨论权重和参数 a 值对新算法优化结果的影响中所使用的患者是同一个人, 所以在新的算法中, 其权重以及各子目标所需参数和前面实验的最初始参数相同。对于基于 DVH 准则的方案优化算法, 每个危及器官的剂量分布由 3 个 DVH 子目标函数 $f_{DVH}(\mathbf{D}(\mathbf{x}))$ 的加权和进行控制, 即用 3 个 DVH 子目标函数 $f_{DVH}(\mathbf{D}(\mathbf{x}))$ 替换公式 3.38 中的一个 NTCP 子目标函数 $f_{NTCP}(\mathbf{D}(\mathbf{x}))$ 。对于前列腺肿瘤病例中的两个危及器官, 直肠和膀胱, 其 DV 约束条件分别为: $V_{50} < 37\%$, $V_{65} < 25\%$ 和 $V_{75} < 15\%$ 与 $V_{65} < 10\%$, $V_{70} < 8.2\%$ 与 $V_{75} < 6\%$; 权重系数分别为 5, 5 和 5 与 5, 8 和 30。将公式 3.38 中控制 PTV 剂量分布的两个子目标函数的权重系数都变为 180 且保持其它参数不变, 就得到这例病人基于 DVH 准则的方案优化算法的总目标函数。图 3-6 为两种方案优化算法的计算结果。我们可以观察到两种优化方法的结果是非常相似的, 如图 3-6(b)所示。但是, 在保证 PTV 剂量覆盖相同的情况下, 新提出的方法可以更好地控制危及器官内高剂量的分布, 如图 3-6(b)所示。与基于 DVH 准则的方案优化算法相比, 新方法可以估算危及器官出现并发症的概率, 而不仅仅是满足 Marks 等人^[56]给出的临床实践指导中的几个 DV 约束条件。在这两个方案优化结果中, 膀胱的 NTCP 值为 18.49% (DVH)和 16.87% (EC-NTCP); 直肠的 NTCP 值为 6.02%(DVH)和 5.80% (EC-NTCP)。相比基于 DVH 准则的方案优化算法, 新算法分别降低了膀胱和直肠出现并发症的概率。

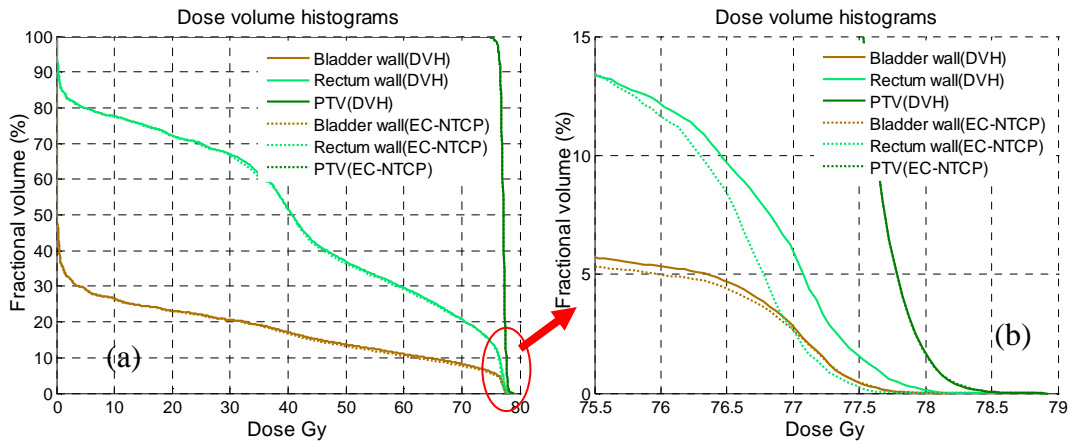


图3-6 新算法和基于DVH准则的方案优化算法在1例前列腺患者上的优化结果。(a) 两种算法优化结果的DVH曲线; (b) 两种算法优化结果的DVH曲线的高剂量部分(图(a)中红色圆圈内的DVH曲线)。

然后, 新算法和基于 DVH 准则的方案优化方法在 10 例前列腺肿瘤患者上进行了比较。由于部分危及器官被 PTV 覆盖, 因此在减少部分危及器官的高剂量分布的同时, 也减少了 PTV 区域内的高剂量分布, 增加了 PTV 区域内剂量分布的非均匀性。Marks 等人^[56]给出的临床实践指导中的各危及器官的 DV 约束条件是在分析大量临床病例的基础上得到的。由于个体(病人体内各器官的大小、相对位置等)之间的差异, 可以适当收紧或放松部分危及器官的 DV 约束, 而且两种优化算法中危及器官的权重系数也应做适当的改变。我们使用试误(trial and error)的方法来确定总目标函数中的权重系数以保证两种优化算法的结果比较相似。在这 10 例前列腺肿瘤患者上, 两种优化算法所使用的权重系数分别如表 3.5 和表 3.6 所示。图 3-7 显示了两种算法在 10 例前列腺肿瘤患者上的平均优化结果。与基于 DVH 准则的方案优化算法相比, 新算法可以更好地控制危及器官的高剂量分布, 如图 3-7(b)所示。美中不足的是新

算法增加了部分低剂量分布(<50Gy)。低剂量分布对有放射治疗而产生的正常组织并发症的影响是有限的。在 Marks 等人^[56]给出的临床实践指导中, 没有给出剂量低于 50 Gy 的 DV 评价指标。

表3.5 10前列腺肿瘤患者在新算法中所使用的权重系数。

	膀胱	直肠	PTV		Tissue ring
	NTCP	NTCP	最小剂量	平均剂量	DVH
病人1	60	90	30	30	5
病人2	45	100	30	30	5
病人3	45	80	30	30	5
病人4	45	110	30	30	5
病人5	50	100	30	30	5
病人6	120	70	30	30	5
病人7	45	80	30	30	5
病人8	45	60	30	30	5
病人9	100	120	30	30	5
病人10	60	50	30	30	5

表3.6 10前列腺肿瘤患者在基于DVH准则的方案优化方法中所使用的权重系数。

	膀胱			直肠			PTV		Tissue ring
	DVH	DVH	DVH	DVH	DVH	DVH	最小剂量	平均剂量	DVH
病人1	5	5	5	5	8	30	180	180	5
病人2	5	5	5	5	8	20	180	180	5
病人3	5	5	5	8	15	15	180	180	5
病人4	8	8	8	12	8	10	180	180	5
病人5	3	3	3	15	15	11	180	180	5
病人6	50	50	50	8	12	10	180	180	5
病人7	1	1	1	8	8	8	180	180	5
病人8	10	10	10	10	10	10	180	180	5
病人9	10	1	1	10	10	10	180	180	5
病人10	10	15	1	10	5	5	180	180	5

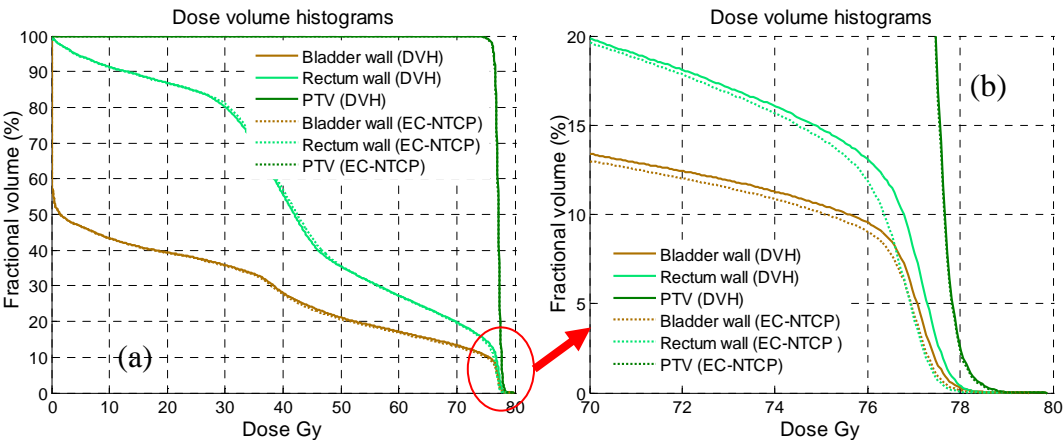


图3-7 新算法和基于DVH准则的方案优化算法在10例前列腺患者上的平均优化结果。(a) 两种算法优化结果的DVH曲线; (b) 两种算法优化结果的DVH曲线的高剂量部分(图(a)中红色圆圈内的DVH曲线)。

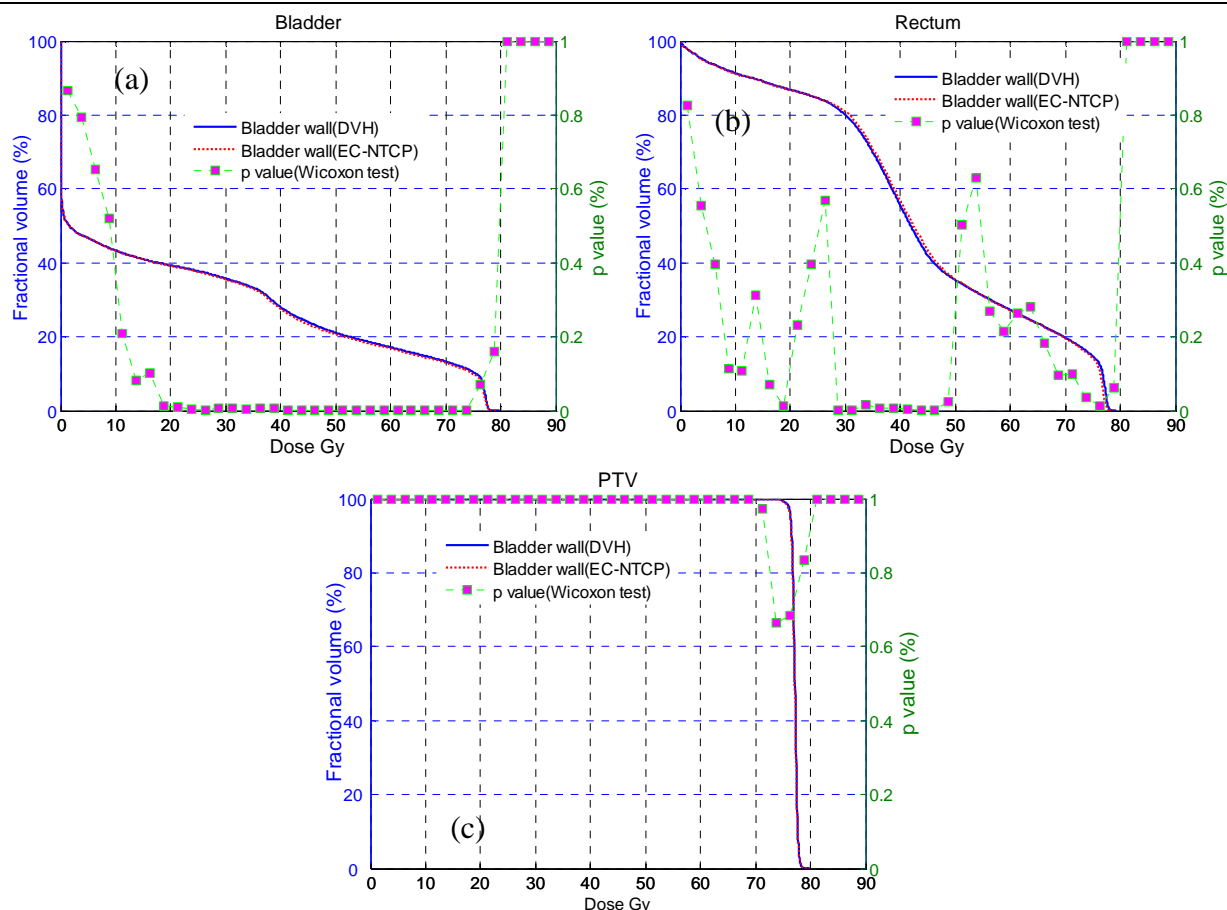
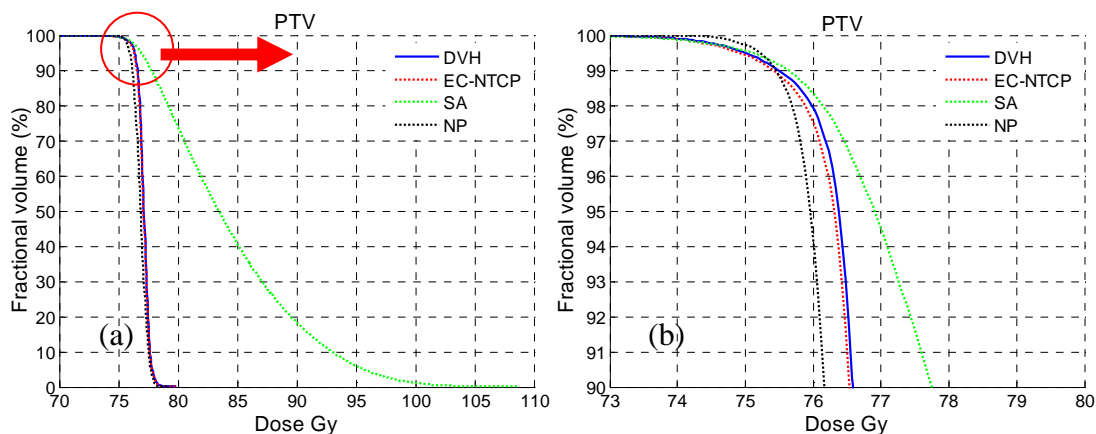


图3-8 图3-7中两组DVH曲线在不同剂量分布范围内的 p 值。(a)膀胱；(b)直肠；(c)PTV。

为了检测两种不同方案优化方法之间的相关性，我们使用 Wilcoxon 检测方法计算图 3-7 中两组 DVH 曲线的相关性，如图 3-8 所示。由于 Wilcoxon 检测方法与使用数据长度有密切的关系，因此我们在不同剂量分布区间内计算 p 的值。从图 3-8 中可知，两种方案优化算法在危及器官上是不相关的。在膀胱内，在大部分的剂量区间内的最大 p 值为 0.0409 ($p < 0.05$)。而在直肠内，两种方案优化方法的差异主要表现在低剂量区域(30-50Gy)和高剂量区域(75-78Gy)。在这两个区域内的最大 p 值为 0.0386 ($p < 0.05$)。而在靶区，两组 DVH 曲线在不同剂量区间的 p 值都大于 0.7737 ($p > 0.05$)，这表明这两组 DVH 曲线的差异是非常小的。这是因为，在本章的工作中，我们在保证靶区覆盖剂量相同的情况下，比较两种方案算法在危及器官(膀胱和直肠)上的差异。因此，在靶区计算得到比较高的 p 值是符合本章的实验设置。



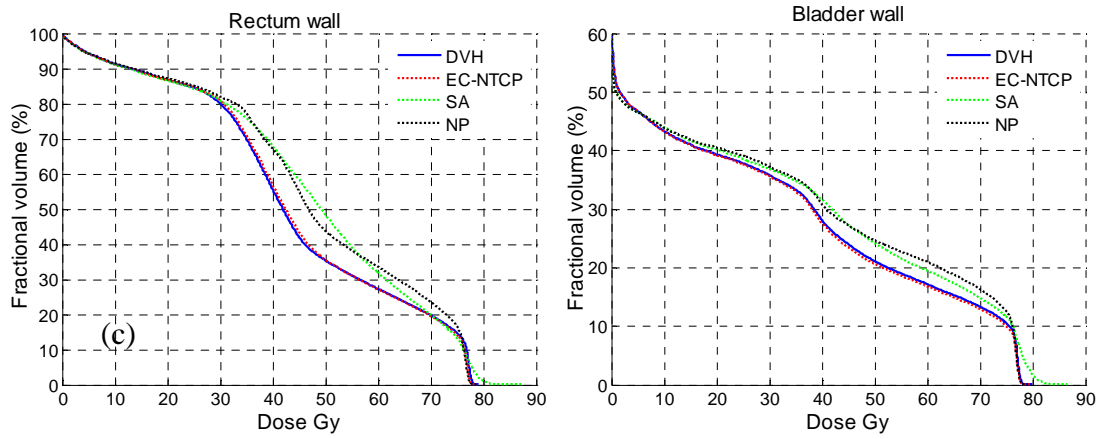


图 3-9 新算法和基于 DVH 准则的方案优化算法以及其它两种基于 NTCP 准则的优化算法在 10 例前列腺患者上的平均优化结果。四种算法优化结果在：(a)靶区的 DVH 曲线；(b)靶区 DVH 曲线的高剂量部分(图(a)中红色圆圈内的 DVH 曲线)；(c)直肠壁的 DVH 曲线；(d)膀胱壁的 DVH 曲线。

3.6.5.3 与其它基于 NTCP 准则的方案优化方法进行比较

本章提出的新算法还分别与 Mohan^[46]等人和 Stavrev 等人^[2]提出的方案优化算法进行了比较，如图 3-9 所示。在 Mohan^[46]等人提出的方案优化算法中，分别使用不同的权重系数表示膀胱、直肠和靶目标在总目标函数中的重要性。而在 Stavrev 等人^[2]提出的方案优化算法中，总目标函数由两 NTCP 模型构成；在其约束条件中，靶区剂量分布分别由最小剂量 76Gy 和最大剂量 79Gy 控制。新算法和基于 DVH 准则的方案优化算法的参数设置分别如表 3.5 和表 3.6 中 10 例患者所用参数。图 3-9 为这四种方案优化算法在 10 例前列腺肿瘤患者上的平均优化结果。与新算法相比，Mohan 等人^[46]提出的方案优化算法对靶区高剂量分布的控制比较差，靶区大约 38% 的剂量分布高于 85.8Gy(110% 的靶区处方剂量)，存在大量热点，靶区剂量均一性很差，如图 3-9(a)所示。由于部分危及器官被 PTV 覆盖，因此 Mohan^[46]等人的算法对靶区高剂量控制力较差同时也影响了危及器官内的高剂量分布(如图 3-9(c)和(d)所示)，直肠和膀胱的最高剂量分别高达 87Gy 和 85Gy。而且，由于直肠和膀胱都是串行器官，对高剂量分布非常敏感，NTCP 值分别增加了 1.23 % 和 6.43%。与新算法相比，Stavrev 等人^[2]等人提出的方案优化算法严格控制了靶区剂量分布，增加了靶区剂量分布的均一性，但却是以减少了对危及器官内高剂量分布的控制为代价，直肠和膀胱的 NTCP 值分别增加了 1.28% 和 3.75%，如图 3-9(b)和(c)所示。四种方案优化算法所得平均结果的 gEUD 值、NTCP 值和 TCP 值如表 3.7 所示。计算靶区的 gEUD 所使用的参数值为 $a=-10^{[13]}$ 。利用 Mohan 等人^[46]在方案优化算法中使用的 TCP 模型及相应参数计算靶区 TCP 值。在危及器官内，利用新算法所得优化结果的 gEUD 值和 NTCP 值要较其它三种方案优化算法低，更好的保护了危及器官。所有的四种方案优化算法都是使用同一台电脑(Intel(R) Xeon(R) 4CPU E5430 平台，4GB 内存)在 MATLAB 平台上实现的。我们还比较了这四种方案优化算法的迭代次数、计算时间和迭代结束后总目标函数的值，如表 3.8 所示。从表 3.8 中可以观察到，新算法和基于 DVH 准则的方案优化算法所需迭代次数、计算时间以及最终总目标函数的值都比其它两种优化算法少。新算法和基于 DVH 准则的方案优化算法都是利用 L-BFGS 算法求解优化问题。因此，使用梯度类的优化算法比随机类的优化算法(模拟退火算法和非线性规划算法)优化效率更高。在基于 DVH 准则的方案优化算法中，每个危及器官的剂量分布由 3 个 DVH 子目标函数控制。由于子目标函数个数的增加，增加了总目标函数的复杂度。而在新算法中，使用复杂的 NTCP 子目标函数(公式 3.36)及其复杂的一阶导数(公式 3.37)，从而增加总目标函数的复杂度。这两种方法相比，新算法复杂度更大，需要更长的计算时间。

表3.7 新算法和基于DVH准则的方案优化算法以及其它两种基于NTCP准则的优化算法在10例前列腺患者上所得平均优化结果的gEUD值、NTCP值和TCP值。

	膀胱		直肠		PTV	
	gEUD (Gy)	NTCP	gEUD (Gy)	NTCP	gEUD (Gy)	TCP
DVH	62.07	51.89%	63.03	6.12%	77.06	53.77%
EC-NTCP	61.76	50.65%	62.78	5.85%	77.04	53.70%
SA	63.27	57.08%	63.70	7.08%	82.67	68.44%
NP	62.57	54.40%	63.98	7.13%	76.80	53.10%

表3.8新算法和基于DVH准则的方案优化算法以及其它两种基于NTCP准则的优化算法的迭代次数、运行时间以及最终目标函数值。

	DVH	NTCP	SA	NP
迭代次数	39	44	10000	99
运行时间	96.27s	98.66s	166.43s	799.63s
目标函数值	0.1146	0.1064	0.3971	0.1409

3.6.6 讨论

在这一小节中我们构造了一个基于物理-生物准则的混合模型的逆向计划方案优化算法。在新算法中，每个危及器官的剂量分布由一个 NTCP 子目标函数控制，而靶区剂量分布则由基于物理准则的子目标函数构成。由于新算法的总目标函数是凸函数，因此使用快速的基于梯度类的算法，L-BFGS 算法，解大规模受限最小化优化问题。

为了验证新算法的有效性，我们在 10 例前列腺肿瘤患者上与基于 DVH 准则的方案优化算法做了对比试验。两种算法在一个病人上的优化结果如图 3-6 所示，在 10 个病人上的平均计算结果如表 3.7 所示。从这两组数据中可以观察到，两种优化算法的优化结果非常相似。这主要是由以下三个原因造成的：1) 我们使用相同的物理准则^[56]来评价方案优化结果。2) 部分危及器官被 PTV 覆盖，过分减少危及器官的高剂量分布，也会同时影响到靶区剂量的覆盖。因此，本章中所有的实验是在保证靶区剂量覆盖尽可能相同的情况下进行的。3) 基于 LQ-泊松模型的 NTCP 模型是关于 gEUD 的 S 形函数，而 gEUD 是一个“特殊的”DV 约束。因此，新算法的优化结果可以被基于 DVH 准则的方案优化算法再实现。每个危及器官上使用越多的 DVH 子函数控制该器官的剂量分布，两种方案优化方法所得优化结果越相近。但是，使用基于 NTCP 准则的优化算法，在优化过程中有估算正常组织出现并发症的概率的潜力，从而能够得到与临床实践密切相关的、最优的生物优化结果。在将来，生物准则将逐步取代物理准则被用来评价方案优化结果^[57]。在这种情况下，基于 DVH 准则的方案优化算法仍然可以得到有效的优化结果。但是，由于基于 DVH 准则的方案优化算法不能直接为当前得到计划提供一个可靠的评价，因此在当前优化结束后，还有必要使用更加严格的 DV 约束条件进行下一次优化，尝试得到更好的优化结果，直到没有更好的优化结果为止。例如，在基于 DVH 准则的优化算法中，开始只能在随机给出的 DV 约束条件下进行优化，并在该优化结果的基础上，根据 DVH 曲线分布情况调整 DV 约束条件，期望在下次优化过程中能得到更好的结果。在整个优化过程中，没有估算危及器官出现并发症的概率，只能继续尝试不同的 DV 约束条件，直到没有更优的解。这种方法也可能得到最优解，但相比基于 NTCP 准则的方案优化算法，需要更多的工作量去完成一个计划。此外，使用基于 NTCP 准则的方案优化算法减少了子目标函数的个数，因而减少了因选取各子目标函数权重所需的时间。

本节还讨论了新算法的 NTCP 子目标函数中使用不同的权重及参数 a 对优化结果的影响。增加 NTCP 子目标函数的权重，目标函数给对相应危及器官的惩罚将增多，因此照射到该危机器官的剂量将会减少。但是权重系数太大，又影响到了靶区剂量分布的均一性，从而影响到放射治疗的质量。因此，子目标函数权重的选取应该综合地考虑靶区 TCP 的减小及危及器

官 NTCP 的降低。参数 a 的值控制着危及器官内主要剂量减少的区域。随着 a 值的增加, 主要剂量减少区域由低剂量区域逐渐向高剂量区域移动。

在使用 NTCP 模型之前, 必须先对所有参数进行校正, 随机选取参数将会使 NTCP 模型失去生物意义。幸运的是, 现在已有大量文献报道 NTCP 模型的参数选取^[13, 53, 54, 58-64]。我们将 Fiorino 在文献^[60]中给出的 6 组不同的参数分别应用到新算法, 并求解在 10 例前列腺肿瘤患者上的优化结果和统计直肠壁的 NTCP 值。基于 LQ-泊松模型的 NTCP 模型在这 6 组模型参数下相对于 gEUD 的输出如图 3-10 所示。我们分别使用 Fiorino 等人^[60]、Rancati 等人^[61]、Soehn 等人^[62]和 Tucker 等人^[63]提出的不同参数计算 RTOG 等级为 2-3 的直肠的 NTCP 值; 分别使用 Rancati 等人^[60]和 Peeters 等人^[64]给出的不同参数计算 RTOG 等级为 3 的直肠的 NTCP 值。其中, Peeters 等人^[64]给出的参数为前面实验所用参数。使用这 6 组参数得到的 10 位患者的直肠 NTCP 值如图 3-11 所示。从图中可以观察到, Peeters 等人^[64]给出参数的优化结果与其余 5 组参数的优化结果有明显的差别, 除 Rancati 等人^[60]提出的 RTOG 等级为 2-3 的直肠的参数之外, 其余四组参数的优化结果中直肠 NTCP 值的 p 值均小于 0.0046 ($p < 0.05$)。其中较大的 p 值可能是由于所使用病例数过少造成的。由此可见, 使用不同参数得到的优化结果相互之间是不相关的。

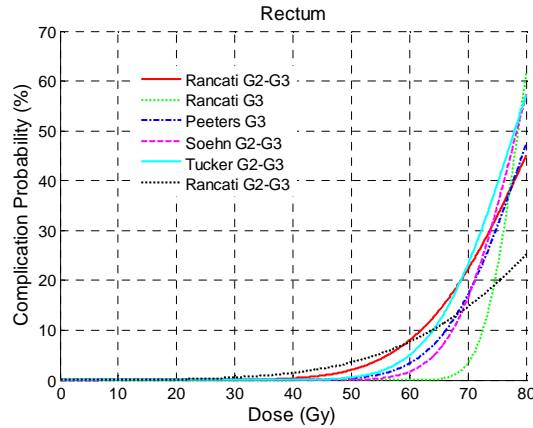


图3-10 基于LKB模型的NTCP准则在不同的模型参数下相对于gEUD的输出。

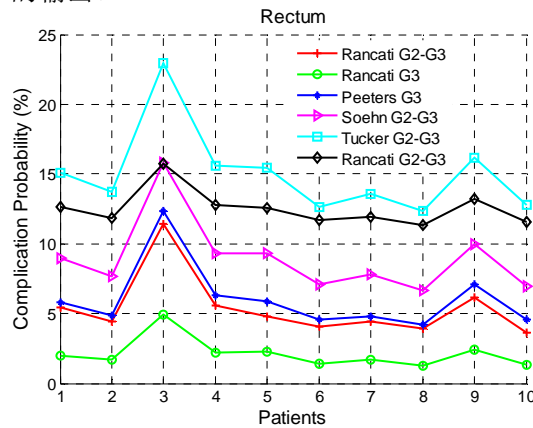


图3-11 在新算法中, 使用不同NTCP模型参数计算10例前列腺患者优化结果中直肠的NTCP值。

本章提出的新算法还与文献[2]和文献[46]中提到的基于 NTCP 准则的方案优化算法进行了比较。如图 3-9 所示, 新算法所得方案优化质量要高于其它两种基于 NTCP 准则的方案优化算法。Mohan 等人^[46]的方法对靶区高剂量分布的控制比较差, 减少了靶区剂量分布的均一性, 影响了放射治疗质量。而且由于部分危及器官包含在 PTV 区域内, 从而导致了危及器官内的高剂量分布较难控制。Stavrev 等人^[2]的方法增加了靶区剂量分布的均一性, 但减少了对危及器官高剂量分布的控制。对于危及器官, 直肠壁, Mohan 等人^[46]方法的优化结果的最高

剂量要大于 Stavrev 等人^[2]方法的,而 Mohan 等人^[46]方法的优化结果的 NTCP 值要小于 Stavrev 等人^[2]方法的。究其原因是, Mohan 等人^[46]的方法在增加了部分高剂量分布的同时减少大部分次高剂量。但是,直肠壁是串行器官,对高剂量特别敏感。因此在使用当前参数的情况下,NTCP 模型不一定能准确的估算危及器官出现并发症的概率。同样,这两种基于 NTCP 准则的方案优化算法也被用来计算 10 例前列腺肿瘤患者最优注量图分布。我们记录了前面介绍的 4 种方案优化算法在这 10 例患者上进行优化所需时间,如图 3-12 所示。我们可以观察到,相比模拟退火算法和非线性规划算法,基于梯度的优化算法需要较少的计算时间,这是由于基于梯度类的优化算法有较高的收敛率。相对于本章提出的算法,模拟退火算法和非线性规划算法关于计算时间的 p 值分别为 0.014 和 0.002 ($p < 0.05$)。较小的 p 值说明这两类优化算法所需计算时间和基于梯度的算法所需计算时间是不相关的,进一步说明使用基于梯度的算法所节省的时间是非常明显的。

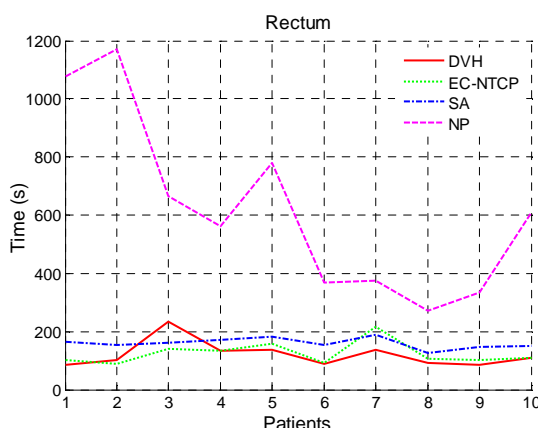


图 3-12 新算法和基于 DVH 准则的方案优化算法以及其它两种基于 NTCP 准则的优化算法在 10 例前列腺患者上进行优化所需的计算时间。

3.7 本章小结

在本章中,我们首先介绍了一些常用的物理准则模型和生物准则模型,然后,介绍了使用基于生物优化准则的方案优化算法的优点,并介绍了将这些准则转换为等效凸准则的方法及部分转换结果。随后,我们提出了本章所使用方案优化算法的总目标函数构成方法以及相应子目标函数的构造方法。最后,我们在前列腺肿瘤患者上验证了新算法的有效性。

在本章中,我们提出了一种将等效凸生物准则应用到方法优化算法中的方法。利用等效凸生物准则构造子目标函数并给出其一阶导数。利用基于梯度的优化算法解 FMO 问题,并与基于 DVH 准则的方案优化算法在 10 例前列腺肿瘤患者上做了对比实验。在对比实验中,新算法利用 NTCP 子目标函数控制危及器官剂量分布,而使用基于物理准则的子目标函数控制靶区剂量分布的均一性。结果表明,可以利用新算法得到临床可接受的方案优化结果。基于 NTCP 准则的方案优化算法能够直接估算危及器官发生并发症的概率,与临床实解释切相关,因此该算法能够提供对整体治疗效果有利的方案优化结果。新算法还与其它两种基于 NTCP 准则的方案优化算法进行了比较。这两种方案优化算法分别使用模拟退火方法和非线性规划方法求解 FMO 问题。对比实验表明,新方法可以得到更高质量的治疗计划,同时其所需计算时间也比其它两种基于 NTCP 准则的方案优化算法少。在将来的工作中,我们将把这种基于物理-生物准则的混合模型的优化算法应用到直接子野优化算法中,进一步提高方案优化的质量,尤其是在使用 column-generation 的直接子野优化算法中^[65]。这类方案优化算法也是利用预先计算好的剂量效应矩阵进行方案优化,因此具有较好的可行性。

3.8 参考文献

[1] Wu Q, Mohan R, Niemierko A, et al. Optimization of intensity-modulated radiotherapy plans

- based on the equivalent uniform dose [J]. *Int J Radiat Oncol Biol Phys*, 2002, 52(1):224-235.
- [2] Stavrev P, Hristov D, Warkentin B, et al. Inverse treatment planning by physically constrained minimization of a biological objective function [J]. *Med Phys*, 2003, 30(11):2948-2958.
- [3] Hindi H A tutorial on convex optimization [C]. *Proc. of the American Control Conf*, Boston: IEEE, 2004, 4: pp 3252–65
- [4] Nesterov Y and Nemirovski A Interior-point polynomial methods in convex programming [M]. Philadelphia, PA: SIAM, 1994, pp. 217-248. (SIAM Series in Applied Mathematics)
- [5] Romeijn HE, Dempsey JF, Li JG. A unifying framework for multi-criteria fluence map optimization models [J]. *Phys Med Biol*, 2004, 49:1991-2013.
- [6] Craft DL, Halabi TF, Shih HA, et al. Approximating convex pareto surfaces in multiobjective radiotherapy planning [J]. *Med Phys*, 2006, 33(9):3399-3407.
- [7] Hoffmann AL, Siem AY, den Hertog D, et al. Derivative-free generation and interpolation of convex Pareto optimal IMRT plans [J]. *Phys Med Biol*, 2006, 51(24):6349-6369.
- [8] Hoffmann AL, Den Hertog D, Siem AY, et al. Convex reformulation of biologically-based multi- criteria intensity-modulated radiation therapy optimization including fractionation effects [J]. *Phys Med Biol*, 2008, 53: 6345-6362.
- [9] Liu DC and Nocedal J. On the limited memory BFGS method for large scale optimization [J]. *Mathematical Programming*, 1989, 45(3):503-528.
- [10] Marks LB, Bentzen SM, Deasy JO, et al. Radiation dose-volume effects in the lung [J]. *Int J Radiat Oncol Biol Phys*, 2010, 76(3 Suppl):S70–S76.
- [11] Deasy JO. Multiple local minima in radiotherapy optimization problems with dose-volume constraints [J]. *Med Phys*, 1997, 24:1157–1161.
- [12] Wu Q and Mohan R. Multiple local minima in IMRT optimization based on dose-volume criteria [J]. *Med Phys*, 2002, 29:1514–1527.
- [13] Li AX, Alber M, Deasy JO, et al. The use and QA of biologically related models for treatment planning: short report of the TG-166 of the therapy physics committee of the AAPM [J]. *Med Phys*, 2012, 39(3):1386-409.
- [14] Marks LB, Yorke ED, Jackson A, et al. Use of normal tissue complication probability models in the clinic [J]. *Int J Radiat Oncol Biol Phys*, 2010, 76(3 Suppl):S10-S19.
- [15] Langer M, Leong J. Optimization of beam weights under dose-volume restrictions [J]. *Int J Radiat Oncol Biol Phys*, 1987, 13(8):1255-1260.
- [16] Niemierko A. Reporting and analyzing dose distributions: a concept of equivalent uniform dose [J]. *Med Phys*, 1997, 24(1):103-110.
- [17] Kutcher GJ, Burman C. Calculation of complication probability factors for non-uniform normal tissue irradiation: the effective volume method [J]. *Int J Radiat Oncol Biol Phys*, 1989, 16(6): 1623-1630.
- [18] Niemierko A. A Generalized Concept of Equivalent Uniform Dose (EUD) (abstract) [J]. *Med Phys*, 1999, 26(6):11110.
- [19] Schultheiss TE, Orton CG, Peck RA. Models in radiotherapy: volume effects [J]. *Med Phys*, 1983, 10(4):410-415.
- [20] Niemierko A. A unified model of tissue response to radiation (abstract) [J]. *Med Phys*, 1999, 26(6): 1100.
- [21] Withers H R and McBride W H 1998 Biologic basis of radiation therapy Principles and Practice of Radiotherapy ed C A Perez and L W Brady (Philadelphia, PA: Williams and Wilkins) pp 79–118 chapter 2
- [22] Stavreva NA, Stavrev PV, Warkentin B, et al. Investigating the effect of cell repopulation on

- the tumor response to fractionated external radiotherapy [J]. *Med Phys*, 2003, 30(5):735-742.
- [23] Fowler JF. The linear-quadratic formula and progress in fractionated radiotherapy. *Br J Radiol*, 1989, 62(740):679-694
- [24] Hall EJ, Giaccia AJ. *Radiobiology for the radiologist* [M]. 6th ed. Philadelphia: Lippincott Williams & Wilkins, 2006, pp. 485-546.
- [25] Zaider M, Minerbo GN. Tumour control probability: a formulation applicable to any temporal protocol of dose delivery [J]. *Phys Med Biol*, 2000, 45(2):279-293.
- [26] Stavreva NA, Warkentin B, Stavrev PV, et al. Investigating the effect of clonogen resensitization on the tumor response to fractionated external radiotherapy [J]. *Med Phys*. 2005, 32(3):720-725.
- [27] Tucker SL, Taylor JM. Improved models of tumour cure [J]. *Int J Radiat Biol*, 1996, 70(5):539-553.
- [28] Kendal WS. A closed-form description of tumour control with fractionated radiotherapy and repopulation [J]. *Int J Radiat Biol*, 1998, 73(2):207-210.
- [29] Sachs RK, Heidenreich WF, Brenner DJ. Dose timing in tumor radiotherapy: considerations of cell number stochasticity [J]. *Math Biosci*, 1996, 138(2):131-46.
- [30] Zaider M, Wu CS, Minerbo GN. The combined effects of sublethal damage repair, cellular repopulation and redistribution in the mitotic cycle. I. Survival probabilities after exposure to radiation [J]. *Radiat Res*, 1996, 145(4):457-466.
- [31] Lyman JT. Complication probability as assessed from dose-volume histograms [J]. *Radiat Res*, 1985, 8(Suppl):S13-19.
- [32] Kutcher GJ, Burman C. Calculation of complication probability factors for non-uniform normal tissue irradiation: the effective volume method [J]. *Int J Radiat Oncol Biol Phys*, 1989, 16(6):1623-1630.
- [33] Withers HR, Taylor JM, Maciejewski B. Treatment volume and tissue tolerance [J]. *Int J Radiat Oncol Biol Phys*, 1988, 14(4):751-759.
- [34] Niemierko A, Goitein M. Modeling of normal tissue response to radiation: the critical volume model [J]. *Int J Radiat Oncol Biol Phys*, 1993, 25(1):135-145.
- [35] Stavrev P, Stavreva N, Niemierko A, Goitein M. Generalization of a model of tissue response to radiation based on the idea of functional subunits and binomial statistics [J]. *Phys Med Biol*, 2001, 46(5):1501-1518.
- [36] Wu Q, Mohan R. Algorithms and functionality of an intensity modulated radiotherapy optimization system [J]. *Med Phys*, 2000, 27:701-711.
- [37] Wu Q, Mohan R. Multiple local minima in IMRT optimization based on dose-volume criteria [J]. *Med Phys*, 2002, 29(7):1514-1527.
- [38] Llacer J, Deasy JO, Portfeld TR, et al. Absence of multiple local minima effects in intensity modulated optimization with dose-volume constraints [J]. *Phys Med Biol*, 2003, 48(2):183-210.
- [39] Wu Q, Mohan R, Niemierko A, et al. Optimization of intensity-modulated radiotherapy plans based on the equivalent uniform dose [J]. *Int J Radiat Oncol Biol Phys*, 2002, 52(1):224-235.
- [40] Choi B, Deasy JO. The generalized equivalent uniform dose function as a basis for intensity-modulated treatment planning. *Phys Med Biol*, 2002, 47:3579-3589.
- [41] Alber M, Nüsslin F. A representation of an NTCP function for local complication mechanisms [J]. *Phys Med Biol*, 2001, 46(2):439-447.
- [42] Roberts SA, Hendry JH. A realistic closed-form radiobiological model of clinical tumor-control data incorporating intertumor heterogeneity [J]. *Int J Radiat Oncol Biol Phys*, 1998, 41(3):689-699.

- [43] Källman P, Agren A, Brahme A. Tumour and normal tissue responses to fractionated non-uniform dose delivery [J]. *Int J Radiat Biol*, 1992, 62(2):249-262.
- [44] Siem AY, den Hertog D, Hoffmann AL. The effect of transformations on the approximation of univariate (convex) functions with applications to Pareto curves [J]. *Eur J Oper Res*, 2008, 189(2):347-362
- [45] Kessler M, Mcshan D, Epelman M, et al. Costlets: A generalized approach to cost functions for automated optimization of IMRT treatment plans [J]. *Optimization and Engineering*, 2005, 6(4): 421-448.
- [46] Mohan R, Mageras GS, Baldwin B, et al. Clinically relevant optimization of 3-D conformal treatments [J]. *Med Phys*, 1992, 19(4):933-944.
- [47] Wang XH, Mohan R, Jackson A, et al. Optimization of intensity-modulated 3D conformal treatment plans based on biological indices [J]. *Radiother Oncol*, 1995, 37(2):140-152.
- [48] Das S. A role for biological optimization within the current treatment planning paradigm [J]. *Med Phys*, 2009, 36(10):4672-4682.
- [49] Dirscherl T, Alvarez-Moret J, Bogner L. Advantage of biological over physical optimization in prostate cancer [J]? *Z Med Phys*, 2011, 21(3):228-235.
- [50] Diot Q, Kavanagh B, Timmerman R, et al. Biological-based optimization and volumetric modulated arc therapy delivery for stereotactic body radiation therapy [J]. *Med Phys*, 2012, 39(1):237-245.
- [51] Deasy JO, Blanco AI, Clark VH. CERR: a computational environment for radiotherapy research [J]. *Med Phys*, 2003, 30(5):979-985.
- [52] Ahnesjö A, Saxner M, Trepp A. A pencil beam model for photon dose calculation [J]. *Med Phys*, 1992, 19(2):263-273.
- [53] Dale E, Hellebust TP, Skjærnsberg A, et al. Modeling normal tissue complication probability from repetitive computed tomography scans during fractionated high-dose-rate brachytherapy and external beam radiotherapy of the uterine cervix [J]. *Int J Radiat Oncol Biol Phys*, 2000, 47(4):963-971.
- [54] Peeters ST, Hoogeman MS, Heemsbergen WD, et al. Rectal bleeding, fecal incontinence, and high stool frequency after conformal radiotherapy for prostate cancer: normal tissue complication probability modeling [J]. *Int J Radiat Oncol Biol Phys*, 2006, 66(1):11-19.
- [55] Mageras GS, Mohan R. Application of fast simulated annealing to optimization of conformal radiation treatments [J]. *Med Phys*, 1993, 20(3):639-647.
- [56] Marks LB, Yorke ED, Jackson A, et al. Use of normal tissue complication probability models in the clinic [J]. *Int J Radiat Oncol Biol Phys*, 2010, 76(3 Suppl):S10-19.
- [57] Ling CC, Li XA. Over the next decade the success of radiation treatment planning will be judged by the immediate biological response of tumor cells rather than by surrogate measures such as dose maximization and uniformity [J]. *Med Phys*, 2005, 32(7):2189-2192.
- [58] Emami B, Lyman J, Brown A, et al. Tolerance of normal tissue to therapeutic irradiation [J]. *Int J Radiat Oncol Biol Phys*, 1991, 21(1):109-122.
- [59] Burman C, Kutcher GJ, Emami B, et al. Fitting of normal tissue tolerance data to an analytic function [J]. *Int J Radiat Oncol Biol Phys*, 1991, 21(1):123-135.
- [60] Fiorino C, Valdagni R, Rancati T, et al. Dose-volume effects for normal tissues in external radiotherapy: pelvis [J]. *Radiother Oncol*, 2009, 93(2):153-167.
- [61] Rancati T, Fiorino C, Gagliardi G, et al. Fitting late rectal bleeding data using different NTCP models: results from an Italian multi-centric study (AIROPROS0101) [J]. *Radiother Oncol*, 2004, 73(1):21-32.

- [62] Fellin G, Fiorino C, Rancati T, V. et al. Late rectal bleeding after conformal radiotherapy for prostate cancer: NTCP Modeling [J]. *Int J Radiat Oncol Biol Phys*, 2008, 72(1):S332–S333.
- [63] Söhn M, Yan D, Liang J, et al. Incidence of late rectal bleeding in high-dose conformal radiotherapy of prostate cancer using equivalent uniform dose-based and dose-volume-based normal tissue complication probability models [J]. *Int J Radiat Oncol Biol Phys*, 2007, 67(4):1066-1073.
- [64] Tucker SL, Dong L, Bosch WR, et al. Fit of a generalized Lyman Normal tissue complication probability (NTCP) model to grade ≥ 2 late rectal toxicity data from patients treated on protocol RTOG 94-06 [J]. *Int J Radiat Oncol Biol Phys*, 2007, 69(3):S8–S9.
- [65] Men C, Romeijn HE, Taşkin ZC, et al. An exact approach to direct aperture optimization in IMRT treatment planning [J]. *Phys Med Biol*, 2007, 52(24):7333-7352.

第四章 IMRT 中校正随机几何不确定性算法的研究

4.1 引言

利用先进的医学成像技术(CT、MRI 和 PET), 剂量计算方法(3D 叠加算法、蒙特卡罗方法)和精确放射治疗技术(IMRT、SRS 和 IGRT), 可以使处方剂量分布与靶区形状高度适形, 从而将正常组织从高剂量区域分离出来。但是, 由于靶目标的几何不确定性, 病人体内实际接收到的 3D 剂量分布会偏离治疗计划中靶区静止不动时的静态剂量分布。在放射治疗过程中, 主要考虑 3 种阻碍精确放射治疗实施的因素: 摆位误差、器官移位和形变、机械误差^[1]。

几何不确定性(geometric uncertainties)可以被分为随机不确定性和系统不确定性^[2]。利用自适应放射治疗摆位方案^[3]和在分次照射时使用在线图像引导技术(EPID, on-board kV imaging, CBCT), 可以明显地减少系统不确定性。但是, 这些技术不能完全消除由成像设备和放射治疗系统的机械限制以及在治疗过程中病人/器官的运动所引起的随机误差。在基于动态多叶准直器的 IMRT 中, 由于器官运动和准直器叶片运动的相互作用, 使得病人实际接收到的剂量分布为平滑后的静态剂量分布。随机不确定性通常可以被抽象为一个描述器官位置的概率分布函数, 如高斯函数^[4-8]。因此, 平滑后的静态剂量分布等于静态剂量分布和概率密度函数的卷积。由于有限的反应时间及运动范围大小, 实时运动跟踪技术只能减小随机不确定性, 但不能完全消除。本章的主要工作就是减少放射治疗过程中由不同因素引起的随机几何不确定性。

传统的校正几何不确定性的方法是扩展射野范围, 将 CTV 扩展到 PTV 来覆盖所有靶目标可能出现的位置^[9]。这种扩展 CTV 的方法导致了靶区周围正常组织接受到了高剂量的照射, 增加了正常组织出现并发症的概率, 影响了放射治疗的质量, 特别是当危及器官被包含在 CTV 的扩展区域内。此外, 在 3D 适形计划中尽管这种方法保证了射野完全覆盖了运动的靶区, 但是这种方法不能保证靶区内接收到的剂量与治疗计划中的静态剂量分布相同。通常, 放射治疗优化方法需要同时考虑射野扩展和注量图调整两方面来校正几何误差。

在使用均匀注量分布射野的 3D 适形放射治疗中, Lind 等人^[10, 11]利用一个简单的分析函数来调整入射射束的注量分布, 从而保证在存在几何不确定性的情况下, 靶区仍旧能接收到处方剂量。调整后的注量分布被称为鲁棒(robust)的注量分布, 其显著的特点是在射野的边缘位置处增加注量强度。这种方法可以被简单定性地理解为: 在射野的边缘处给予较高的注量分布; 因此, 处于靶目标边缘的组织在射野边缘位置接收到的高剂量分布正好补偿这些器官移动到射野之外接收到的低剂量照射。整体的输出结果为, 不需要对 CTV 进行扩展, 就可以保证在存在几何不确定性的情况下靶区实际接收到的剂量分布与静态剂量分布相同。在存在几何不确定性的情况下, 当靶区边缘的组织在射野边缘的注量突起周围接收到 100% 的处方剂量, 紧邻靶区的正常组织接收到的剂量要小于 100% 的处方剂量, 尽管靶区在所有方向上都有移动。已有学者证明了这种方法的可行性, 并建议使用增加射野边缘注量分布的鲁棒注量图来取代对 CTV 区域的扩展, 尤其是危及器官紧邻 CTV^[10, 11]。

在 IMRT 中, 2D 鲁棒注量图分布通常是将关于几何不确定性的信息包含到 IMRT 的逆向方案优化算法中计算得到^[8, 12-15]。在包含几何不确定性的优化过程中, 目标函数中对体模内某点处剂量的限制变为对该点在体模内不同位置的平均剂量的限制。利用优化算法所得的鲁棒注量图照射包含几何不确定性的靶目标, 靶目标接收到的实际剂量分布与利用静态注量分布照射静止的靶目标所得剂量分布相同。在利用优化算法求鲁棒注量图的方法中, 计算目标函数中的平均剂量是基于病人的轮廓及体内非均匀介质分布可以被忽略的假设上进行的。这种假设适应于前列腺肿瘤患者, 对于其它部位的肿瘤放射治疗可能并不合适, 如头颈部肿瘤放射治疗^[14]。在头颈部肿瘤患者的方案优化过程中, 剂量计算不可以忽略病人的体表轮廓及内部非均匀介质分布。Moore 等人^[8]对计算目标函数中的平均剂量方法进行了改进, 将几何不确定性直接包含在注量分布的计算过程中, 避免了计算平均剂量所引起的误差。将几何不

确定性加入到优化算法中，增加了方案优化的复杂度。

实际上，求解鲁棒的注量图是一个反卷积的问题。已有一些学者利用反卷积算法来求解存在较小强度随机几何不确定性下的鲁棒注量图^[10, 16]和去除有限探测器尺寸对测量数据的影响^[17, 18]。为了避免直接频域滤波所得到的不可行或不平稳的反卷积结果，序列反卷积算法被用来计算反卷积结果^[16-20]。在序列反卷积算法中，反卷积结果只取序列展开的前 4 项之和，从而减少了反卷积结果中的高频成份。序列反卷积算法可以有效地抑制反卷积结果中的信号振动，但是这种抑制效果是比较有限的，反卷积结果中仍然包含有过多的信息，特别是在靶区的边缘位置处。Fan 等人^[16]提出了一种新的序列展开方法来抑制射野边缘处的过高剂量分布。但是，该方法的抑制效果过大，反而导致了射野边缘位置处的剂量分布低于处方剂量。在本章中，我们提出了一种基于直接序列展开算法和频域滤波的混合反卷积算法。为了抑制信号中的高频成份，反卷积结果只取直接序列展开的前 4 项。为了进一步抑制反卷积结果中的高频成份，我们使用巴特沃斯滤波器(Butterworth)对直接序列展开算法的反卷积结果进行滤波。根据不同随机几何不确定性的概率密度函数，调整巴特沃斯滤波器的参数值，使得在考虑随机几何不确定性的情况下体模内实际剂量分布与体模静止情况下的静态剂量分布相同。

在本章中，我们首先介绍了 IMRT 中的几何不确定性及减少几何不确定性的常用方法；然后，详细地描述了新提出反卷积算法的原理；最后，在一个规则 2D 注量图、一个真实的 IMRT 注量图和一例前列腺肿瘤患者上，分别验证了本章提出算法的有效性。

4.2 IMRT 中的几何不确定性

一个放射治疗过程通常由一个计划设计阶段和多个照射阶段组成。在计划阶段中，利用 CT 数据或模拟图像来确定病人体内器官的几何结构，并以这些器官的几何结构为基础设计放射治疗计划。之后的放射治疗都是按照计划阶段制定的计划进行的。然而，在放射治疗过程中，由于病人体内的几何不确定性，导致靶区剂量不足或靶区周围正常组织接收到的剂量过高。在 ICRU 62 号报告中，主要考虑 3 种降低放射治疗质量的几何不确定性：摆位误差、器官移位和形变、机械误差^[1]。摆位误差是指分次放射治疗时由在治疗床上重新确定的病人位置与计划中病人位置的差异所引起的误差。尽管一些方法被用来提高重新摆位的精度，但是摆位误差仍然是不可避免的，不会被完全消除。肿瘤在病人体内的相对位置也在随时间进行变化。例如，在前列腺肿瘤患者体内，由于每天直肠和膀胱的充盈程度不同而引起的肿瘤位置变化。心跳活动和呼吸运动同样会导致分次照射之间肿瘤位置的移动。在现代放射治疗设备中，机械误差的大小一般要比摆位误差和器官运动小很多，如射束面积和机架角度等。

几何不确定性可以被分为系统几何不确定性和随机几何不确定性两类。如果每次分次照射时病人体内器官的几何结构相对于计划阶段病人体内器官的几何结构有平均偏移，我们就认为存在系统几何不确定性。这个平均偏移就表示系统几何不确定性的大小。在各分次照射时，病人体内器官的几何不确定性围绕平均偏移的偏差被称为随机几何不确定性。例如，所有分次照射时，病人在治疗床上的位置与计划设计阶段获取的 CT 数据中病人的位置的差异被认为是服从随机分布，即随机几何不确定性。

在放射治疗中，由几何不确定性引起的误差需要被校正。很明显，对于一个特定的病人，我们无法估算这个病人在未来治疗过程中的几何不确定性。但是，我们可以从已经接受过放射治疗的相似病人中测量得到几何不确定性数据。如 ICRU 所报到的，一组相似病人的系统和随机几何不确定性都可以分别用一个标准偏差来表示。对于系统不确定性，总体标准偏差(Σ)被用来表示一组病人所有分次照射时的几何结构到计划中病人的几何结构的系统偏差。总体标准偏差(σ)用来表示随机几何不确定性，即在一组病人上观察到的随机标准差的平均值。由相互独立的几何不确定性源造成的误差也应该是相互独立的。例如， x -轴方向的平移不确定性可以表示为由内部器官运动引起的内部几何不确定性 $\Sigma_{int,x}$ 和 $\sigma_{int,x}$ ，和由外部摆位误差引起的外部几何不确定性 $\Sigma_{ext,x}$ 和 $\sigma_{ext,x}$ 。从一组相似的病人上计算得到的 Σ -和 σ -则可以被用来估

算一个新病人的几何不确定性。在得到 Σ -和 σ -值之后，还需要一个概率密度函数来近似的描述几何不确定性的运动分布情况，通常被表示为高斯函数^[4-8, 16]。

在本章中，我们只考虑对随机几何不确定性的校正，并在一个规则 2D 注量图、一个真实的 IMRT 射野注量图和一例前列腺肿瘤患者上验证本章提出算法的有效性。在临床实践中，病人体内多数器官的随机几何不确定性的标准差 σ 在 3mm 到 4mm 之间^[4-8, 16]。在本章的 3 个实验中，我们同样使用二维高斯函数来描述几何不确定性，

$$p(x, \sigma_x, y, \sigma_y) = \frac{1}{2\pi\sigma_x\sigma_y} \exp\left(-\frac{(x)^2}{2\sigma_x^2} - \frac{(y)^2}{2\sigma_y^2}\right) \quad (4.1)$$

其中在 x, y 方向没有系统几何不确定性， $\sigma_x=4\text{mm}$ 和 $\sigma_y=3\text{mm}$ 分别表示在 x, y 方向的随机几何不确定性大小，如图 4-1 所示。

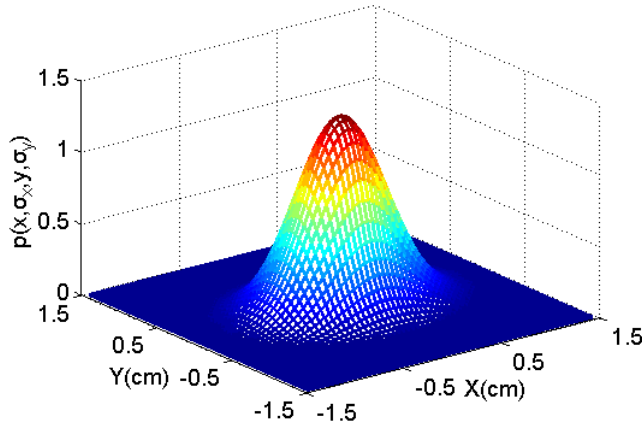


图 4-1 标准差为 $\sigma_x=4\text{mm}$ 和 $\sigma_y=3\text{mm}$ 的高斯分布。

4.3 利用反卷积算法校正 IMRT 中随机几何不确定性



图 4-2 鲁棒注量图、静态注量图 and 实际注量图^[16]。

反卷积是一种常用的数学方法，在图像处理领域经常被用来去除由于物体运动所引起的图像模糊。同样，反卷积算法也被应用于放射治疗领域，如求解存在较小强度随机几何不确定性下的鲁棒注量图^[10, 16]和去除有限探测器尺寸对测量数据的影响^[17, 18]。在治疗计划的设计阶段，方案优化是在该阶段获取的 CT 数据上进行的。此时没有考虑之后分次照射中存在的随机几何不确定性，优化所得注量图被称之为静态注量图。利用静态注量图对包含随机几何不确定性的病人进行照射，病人实际接收到的注量分布相当于病人静止不动，静态注量图按病人的随机不确定性进行运动的情况下病人接收到的注量分布。因此，考虑到病人的随机几何不确定性，利用静态注量图照射病人，病人实际接收的注量分布不再是静态注量图，而是静态注量图和随机几何不确定性的概率密度函数卷积后的模糊注量图，如图 4-2 所示。我们实际想要的是，考虑病人治疗过程中的随机几何不确定性，病人在全部放射治疗之后，实际接收到的注量分布是静态注量分布。也就是说，用一个为未知的注量分布照射病人，由于病人的随机几何不确定性，病人实际接收到的注量分布为静态注量分布。这个未知的注量分布就是我们想要的结果。在已知随机几何不确定性的概率密度函数的情况下，未知注量分布和概率密度函数的卷积为静态注量分布。因此，求解未知的注量图为一个反卷积的过程，如图

4-2 所示。经过计算，我们发现所得未知的注量图具有前面介绍的鲁棒注量图的特征，因此我们也称未知的注量图为鲁棒注量图。

4.3.1 1D 注量图反卷积算法

我们首先介绍 1D 信号中的序列反卷积。在已知 1D 随机几何不确定性的概率密度分布函数 $p(x, \sigma)$ 的情况下，模糊的注量图 \mathbf{D}_c 可以表示为：

$$\mathbf{D}_c = \int \mathbf{D}_0(x') p(x - x', \sigma) dx' \quad (4.2)$$

式中 \mathbf{D}_0 为初始的或静态的注量图。理论上，体模内的几何不确定性属于非刚体运动，每一点 x 处都有一个特定的概率密度函数 $p(x, \sigma)$ 来描述这一点的几何不确定性。在本章中，我们近似地认为靶区的几何不确定性为刚体运动，即靶区内所有点的运动概率密度函数都相同。利用泰勒级数直接将 $\mathbf{D}_0(x')$ 在 x 点展开，则等式 4.2 变为：

$$\mathbf{D}_c = A \mathbf{D}_0 \quad (4.3)$$

其中 $A = \sum_{n=0}^{\infty} \frac{\mu_n}{n!} \frac{d^n}{dx^n}$ ， μ_n 为概率密度函数的 n 阶矩函数：

$$\mu_n = \int (x' - x)^n p(x - x', \sigma) dx' \quad (4.4)$$

算子 A 为序列差分之和，表示式 4.2 中的卷积运算。对于满足高斯分布的 1D 随机几何不确定性概率密度函数：

$$p(x, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{x^2}{2\sigma^2}\right) \quad (4.5)$$

我们可以计算得到高斯函数各阶矩的值，由于高斯函数的对称分布而消除了式 4.4 中的奇数项，算子 A 可以进一步表示为：

$$A = \sum_{n=0}^{\infty} \frac{\sigma^{2n}}{2^n n!} \frac{d^{2n}}{dx^{2n}} = \exp\left(\frac{\sigma^2}{2} \frac{d^2}{dx^2}\right) \quad (4.6)$$

逆向算子 A^{-1} 满足条件： $A^{-1}A = A A^{-1}$ 。当 \mathbf{D}_c 为静态注量图时，可根据逆向算子 A^{-1} 计算得到鲁棒的注量图 \mathbf{D}_0 ：

$$\mathbf{D}_0(x) = A^{-1} \mathbf{D}_c(x) = \exp\left(-\frac{\sigma^2}{2} \frac{d^2}{dx^2}\right) \mathbf{D}_c(x) \quad (4.7)$$

上式介绍的是直接序列展开算法的反卷积结果。为了进一步提高反卷积结果的精度，Ulmer 等人将序列展开算法扩展为反卷积核算法^[17, 18]。由公式

$$\mathbf{D}_0(x) = A^{-1} \mathbf{D}_c(x) = A^{-1} \int \mathbf{D}_0(x') p(x - x', \sigma) dx' = \int \mathbf{D}_0(x') \delta(x - x') dx' \quad (4.8)$$

可知 $A^{-1} p(x, \sigma) = \delta(x)$ 。将该结果代入式 4.7 可得

$$\begin{aligned} A^{-1} \mathbf{D}_c(x) &= \mathbf{D}_0(x) = \int A^{-1} \delta(x - x') \mathbf{D}_c(x') dx' \\ &= \int (A^{-1})^2 p(x - x', \sigma) \mathbf{D}_c(x') dx' \\ &= \int p^{-1}(x - x', \sigma) \mathbf{D}_c(x') dx' \end{aligned} \quad (4.9)$$

式中 $p^{-1}(x, \sigma)$ 为反卷积核，可以进一步将反卷积核展开为 Hermite 多项式 H_{2n} 和的形式：

$$p^{-1}(x, \sigma) = \sum_{n=0}^{\infty} c_n H_{2n} \left(\frac{x}{\sqrt{2\sigma}} \right) p(\sigma, x) \quad (4.10)$$

式中参数 $c_n = (-1)^n / (2^n n!)$ 。

4.3.2 2D 注量图反卷积算法

在本章中，我们利用序列反卷积算法对 2D 注量图进行处理。已知靶目标几何不确定性的 2D 概率密度函数(式 4.1)，且 x 方向和 y 方向的分量相互独立可分离，则可将 1D 直接序列展开算法(4.7)直接扩展为 2D 直接序列展开算法

$$\mathbf{D}_0(x, y) = \exp \left(\frac{\sigma_x^2}{2} \frac{d^2}{dx^2} \right) \exp \left(\frac{\sigma_y^2}{2} \frac{d^2}{dy^2} \right) \mathbf{D}_c(x, y) \quad (4.11)$$

2D 反卷积核算法表示为

$$\mathbf{D}_0(x, y) = \iint p^{-1}(x - x', \sigma_x, y - y', \sigma_y) \mathbf{D}_c(x', y') dx' dy' \quad (4.12)$$

式中 2D 反卷积核

$$p^{-1}(x, \sigma_x, y, \sigma_y) = p^{-1}(x, \sigma_x) p^{-1}(y, \sigma_y) = \sum_{m=0}^{\infty} \sum_{n=0}^{\infty} c_m c_n H_{2m} \left(\frac{x}{\sqrt{2\sigma_x}} \right) H_{2n} \left(\frac{y}{\sqrt{2\sigma_y}} \right) p(x, \sigma_x, y, \sigma_y) \quad (4.13)$$

2D 反卷积核 $p^{-1}(x, \sigma_x, y, \sigma_y)$ 应该还保持着 2D 卷积核 $p(x, \sigma_x, y, \sigma_y)$ 的对称特性。精确地计算 2D 反卷积核需要对所有阶数的 Hermite 多项式 H_{2n} 进行求和。在实际的应用中只取式 4.13 中的前几项进行求和来近似地估算整体的 2D 反卷积核。Fan 等人^[16]利用反卷积核的对称性，对 2D 反卷积方法进行了改进，进一步抑制了反卷积结果中的高频成份。利用公式

$$r = \sqrt{\frac{x^2}{2\sigma_x^2} + \frac{y^2}{2\sigma_y^2}} \quad (4.14)$$

2D 反卷积核(式 4.13)可以改写为：

$$p^{-1}(r, \sigma_x, \sigma_y) = f(r) p(r, \sigma_x, \sigma_y) \quad (4.15)$$

2D 反卷积核只取前 4 阶的 Hermite 多项式 $f(r)$ 可以近似地被表示为：

$$f(r) = \frac{49}{4} - 28r^2 + \frac{53}{4}r^4 - 2r^6 + \frac{3}{32}r^8 \quad (4.16)$$

4.4 改进的反卷积算法

4.4.1 算法描述

在不考虑病人随机几何不确定性的情况下，靶区静止不动，其接收到的静态注量图 (“Origin”) 为均匀注量分布，如图 4-3 所示。在考虑病人随机几何不确定性的情况下，我们希望靶区在运动过程中接收到的注量分布与静态注量分布相同。因此，我们首先利用上面三种反卷积方法分别计算得到鲁棒的注量图。然后，将鲁棒的注量图与病人随机不确定性的概率密度函数卷积，得到靶区在运动过程中接收到的实际注量分布，我们也称之为卷积后注量图。最后，通过对卷积后的注量图与静态注量图进行比较来评价不同反卷积算法的性能。我们分别计算直接序列展开算法(式 4.7)和反卷积核算法(式 4.9)两种序列反卷积方法在 1D 信号中的反卷积结果。反卷积结果与随机不确定性的概率密度函数的卷积后注量分布依次标识为 “Series” 和 “DK”，如图 4-3 所示。从图中我们可以发现，相比直接序列展开方法，反卷积核算法的卷积后注量图更接近静态注量图。因此，反卷积核算法要优于直接序列展开反卷

积算法。

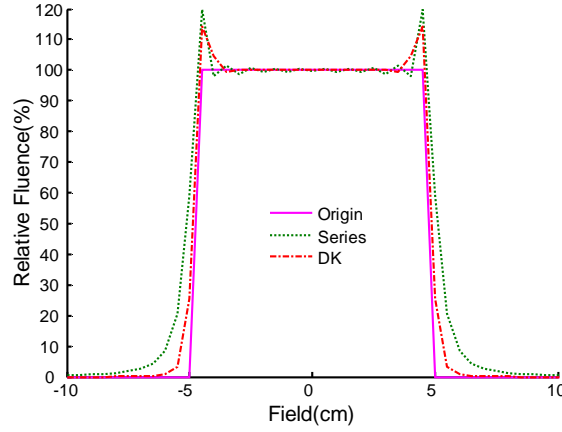


图 4-3 静态注量图和利用直接序列展开方法和反卷积核算法计算得到的卷积后注量分布。

由上面计算结果可知，反卷积核算法的卷积后注量图更接近于静态注量图。我们在频域中对直接序列展开算法和反卷积核算法进行分析。直接序列展开算法(式 4.7)的傅立叶变换展开式为：

$$F\{A^{-1}\mathbf{D}_c(x)\} = \sum_{n=0}^{\infty} \frac{1}{n!} \left(\frac{\sigma^2}{2} \right)^n w^{2n} F\{\mathbf{D}_c(x)\} \quad (4.17)$$

式中 w 为角频率。由式 4.9 可知，反卷积核算法的卷积核还可以表示为：

$$p^{-1}(x, \sigma) = (A^{-1})^2 p(x, \sigma) \quad (4.18)$$

式中 $(A^{-1})^2 = \left(\exp\left(-\frac{\sigma^2}{2} \frac{d^{(2)}}{dx^{(2)}}\right) \right)^2 = \exp\left(-\sigma^2 \frac{d^{(2)}}{dx^{(2)}}\right) = \sum_{n=0}^{\infty} \frac{(-1)^n}{n!} \sigma^{2n} \frac{d^{(2n)}}{dx^{(2n)}}$ 。则在反卷积核算法中卷积核的傅立叶展开式为：

$$F\{p^{-1}(x, \sigma)\} = \sum_{n=0}^{\infty} \frac{1}{n!} \sigma^{2n} w^{2n} \exp\left(-\frac{w^2 \sigma^2}{2}\right) \quad (4.19)$$

因此，反卷积核算法的傅立叶展开式可表示为：

$$F\left\{\int p^{-1}(x-x')\mathbf{D}_c(x')dx'\right\} = \sum_{n=0}^{\infty} \frac{1}{n!} \sigma^{2n} w^{2n} \exp(-w^2 \sigma^2 / 2) F\{\mathbf{D}_c(u)\} \quad (4.20)$$

比较公式 4.17 和公式 4.20 可知，反卷积核算法相当于一个直接序列展开算法和一个滤波器 $H(w)$ 的乘积：

$$F\left\{\int p^{-1}(x-x')\mathbf{D}_c(x')dx'\right\} = H(w) F\{A^{-1}\mathbf{D}_c(x)\} \quad (4.21)$$

式中

$$H(w) = \frac{\sum_{n=0}^{\infty} \frac{1}{n!} \sigma^{2n} w^{2n}}{\sum_{n=0}^{\infty} \frac{1}{n!} \left(\frac{\sigma^2}{2} \right)^n w^{2n}} \exp\left(-\frac{\sigma^2 w^2}{2}\right) \quad (4.22)$$

当 n 趋于无穷大时， $H(w)=1$ 为一个全通滤波器。序列展开式高阶项的贡献相对比较小，且高阶项中包含的高频成份较多。为了能在计算机上进行运算，且减少反卷积结果中的高频分量，

我们只取公式 4.22 中序列展开式的前 4 项进行求和。此时，滤波器的输出相当于一个低通滤波器，如图 4-4 所示。从图中我们可以观察到，滤波器 $H(w)$ 为一个抑制信号高频成份的滤波器。因此，反卷积核算法的计算结果等于对直接序列展开算法的计算结果使用滤波器 $H(w)$ 进行滤波，抑制了信号中的高频成份，特别是在射野的边缘处。所以，反卷积核算法要优于直接序列展开算法。但是，滤波器 $H(w)$ 的频率响应特性只与所使用的序列展开式截断项有关。对于使用固定的序列展开截断项，滤波器 $H(w)$ 对高频成份的抑制作用也是有限的。例如在图 4-3 中，反卷积核算法的卷积后注量分布在射野边缘处与静态注量分布的差值仍然很大。为了进一步缩小这个差值，需要调整滤波器 $H(w)$ 频率响应特性，使其能进一步抑制序列展开式中的高频成份。于是，我们利用经典的低通滤波器代替前面的滤波器 $H(w)$ 。对于经典的低通滤波器，我们可以调整其参数来改变其频率输出特性。在本章中，我们使用巴特沃斯滤波器进行滤波，其频率响应为：

$$H(w) = \sqrt{\frac{1}{1 + (w/w_0)^{2m}}} \quad (4.23)$$

式中 w_0 为截止频率， m 为滤波器的阶数。

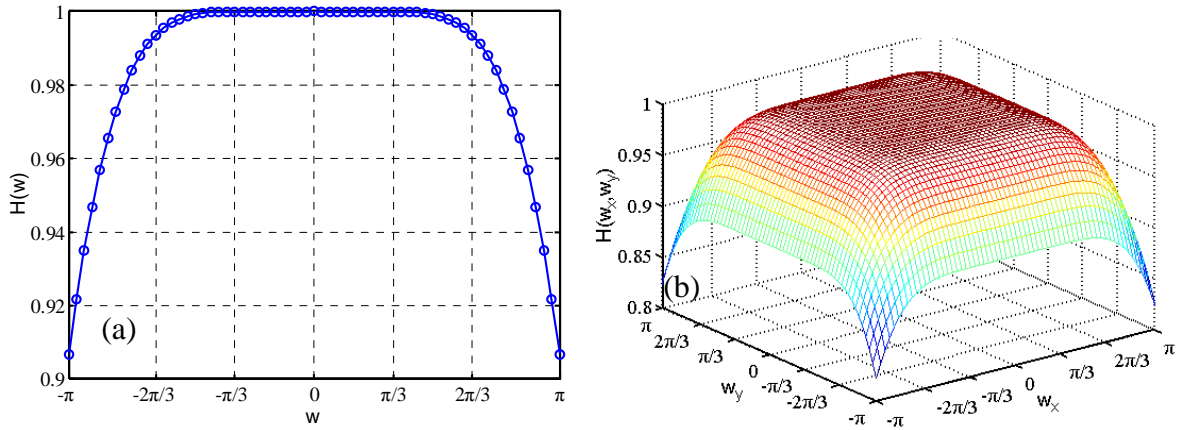


图 4-4 滤波器 $H(w)$ 的频率响应。(a) 1D 频率响应；(b) 2D 频率响应。

在本章中，我们首先根据随机几何不确定性的概率密度函数，利用直接序列展开算法对静态注量图求反卷积。其中直接序列展开算法取前 4 项进行相加，因为序列展开的低阶项包含了信号大部分的能量。然后，利用巴特沃斯滤波器对反卷积结果进行滤波。对于不同的概率密度函数，通过调整滤波器的参数，可以很容易地抑制反卷积信号中的高频成份，并缩小卷积后注量图和静态注量图之间的差值，如图 4-5(a)所示。图中，“Deconvolution”表示利用新方法进行反卷积后所得的鲁棒注量图，“Convolution-back”表示反卷积所得鲁棒注量图与概率密度函数进行卷积所得卷积后的注量图。从图中我们可以观察到，新方法减小了卷积后注量图与静态注量图之间的差值，特边是在射野的边缘处，减小尤为明显。为了进一步减少靶区之外正常组织接收到的剂量照射，我们将射野之外的注量分布设为 0，如图 4-5(b)所示。但是，当靶区边缘的组织运动到射野之外，就接收不到剂量的照射，这种近似方法又减少了靶区边缘组织所接收到的注量分布。这时，我们需要调整滤波器的参数，增加鲁棒注量图在靶区边缘位置的注量，来补偿靶区组织运动到射野之外接收不到射线的照射。我们系统的搜索滤波器参数空间来确定最优的滤波器参数。最优的滤波器参数为卷积后注量图和静态注量图之差最小时所使用的参数。

在 2D 注量图反卷积中，直接序列展开算法(式 4.11)的傅立叶变换为：

$$F(D_0) = \sum_{n=0}^{\infty} \frac{1}{n!} \left(\frac{\sigma_x^2}{2} \right)^n u^{2n} \sum_{k=0}^{\infty} \frac{1}{k!} \left(\frac{\sigma_y^2}{2} \right)^k v^{2k} F\{D_c(x, y)\} \quad (4.24)$$

式中 u 和 v 为角频率。为了抑制反卷积信号中的高频成份，我们利用 2D 巴特沃斯滤波器对反卷积结果进行滤波。2D 巴特沃斯滤波器的频率响应为：

$$H(u, v) = \sqrt{\frac{1}{1 + (u/u_0)^{2m}}} \sqrt{\frac{1}{1 + (v/v_0)^{2k}}} \quad (4.25)$$

式中 u_0 和 v_0 为截止频率， m 和 k 为滤波器的阶数。

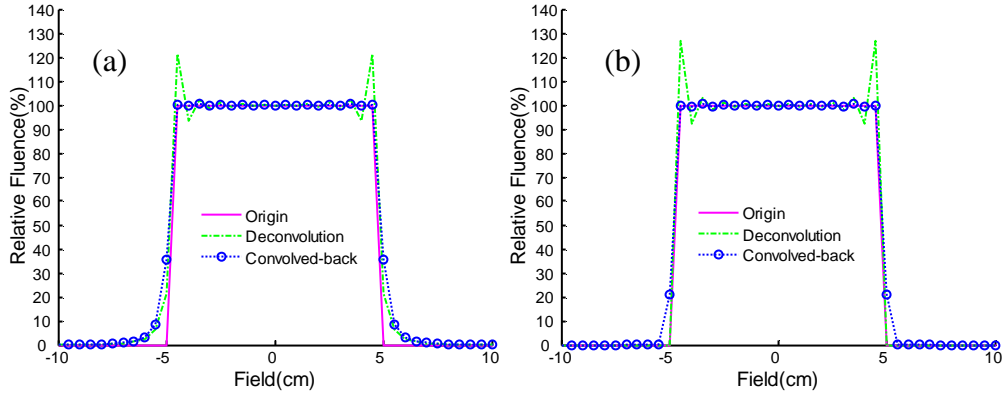


图 4-5 利用新方法所得反卷积结果。(a)新算法；(b)设置射野外注量分布为 0 所得反卷积结果。

本章提出的新方法还与 Fan 等人^[16]提出的改进的反卷积核算法进行了比较，如图 4-6 所示。本文提出算法的计算结果被标识为“Filter”，Fan 等人^[16]的算法的计算结果被标识为“IDK”。反卷积核算法、改进的反卷积核算法和新算法分别计算所得卷积后注量图与静态注量图的差值依次被标识为“DDK”、“DIDK”和“DFilter”。我们从图中可以观察到，改进的反卷积核算法抑制了信号中的高频成份，特别是在射野边缘位置处。但是，这种方法对高频成份的抑制程度过大，导致了在卷积后注量图中靶区边缘位置的剂量不足。

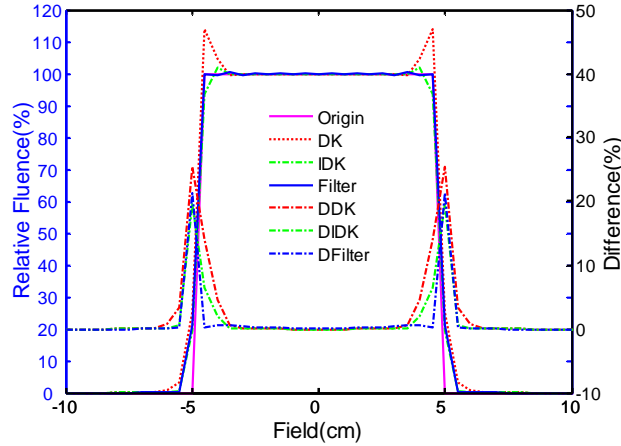


图 4-6 利用反卷积核算法(DK)、改进的反卷积算法(IDK)和新算法(Filter)所得卷积后的注量分布。

4.4.2 实验结果

本章中，我们在两个 2D 注量图(如图 4-7(a)和 4-8(a))和一个前列腺肿瘤患者(图 4-10)上验证新算法的有效性。根据已有文献^[4-8, 16]的介绍，病人的随机几何不确定性可以用高斯函数来描述。在本章，我们也使用高斯函数描述病人随机几何不确定性的概率密度函数。2D 高斯函数在 x 和 y 方向的标准差分别为 $\sigma_x=4\text{mm}$ 和 $\sigma_y=3\text{mm}$ 。为了抑制直接序列展开算法反卷结果中的高频成份，我们需要利用一个高截止频率、低阶数的 2D 巴特沃斯滤波器(式 4.25)对反卷

积结果进行滤波，其参数分别为 $u_0 = 0.9375$, $m = 1$ 和 $v_0 = 1.25$, $k = 1$ 。截止频率 u_0 和 v_0 是利用穷举算法计算得到的。最优截止频率为卷积后注量图和静态注量图之差最小时所使用的截止频率。截止频率只与高斯函数的标准差有关。对于相同大小的几何不确定性，我在不同的注量图反卷积过程中使用相同的截止频率。

我们计算卷积后注量图的剂量分布，进一步来验证反卷积算法的有效性。剂量计算的步骤如下：

- (1)在已知静态注量图的情况下，利用反卷积算法计算鲁棒的注量图。
- (2)将鲁棒的注量图分割为一系列机器可执行的多叶准直器叶片序列，然后用这些可执行的叶片序列重构出可执行的注量图。
- (3)将可执行的注量图与表示病人随机几何不确定性的高斯函数进行卷积，得到卷积后注量图。
- (4)最后，在放射治疗开源软件 CERR^[21]上计算卷积后注量图的剂量分布。

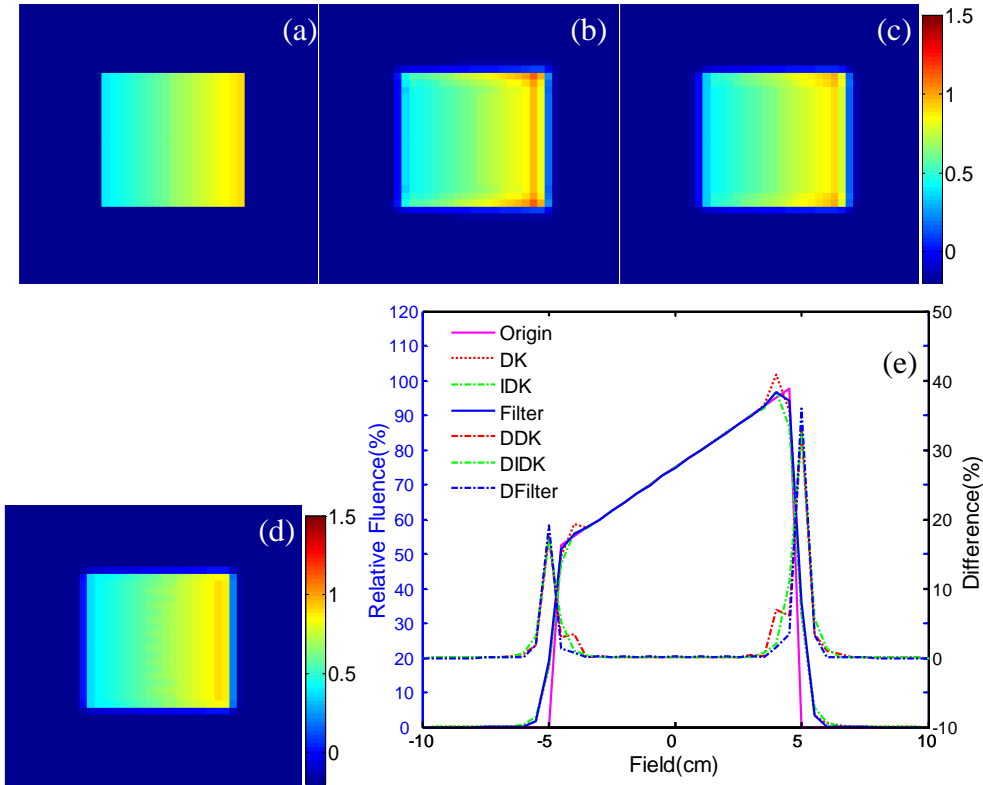


图 4-7 静态注量图和卷积后注量图及其剖面曲线。(a)静态注量图；(b)反卷积核算法的卷积后注量图；(c) 改进的反卷积核算法的卷积后注量图；(d) 新算法的卷积后注量图；(e) 静态注量图和卷积后注量图的剖面曲线。

4.4.2.1 规则 2D 注量图

图 4-7 为利用序列反卷积算法对规则的 2D 注量图进行反卷积所得鲁棒注量图的卷积后注量图。在理想情况下，卷积后注量图(图 4-7(b)-(d))应该与实际注量图(图 4-7(a))相同。而在实际中，两者是有一定差别的。由于静态注量图强度在射野边缘陡峭的下降，其差别主要集中在射野的边缘位置处。图 4-7(e)为静态注量图和卷积后注量图的剖面曲线比较。

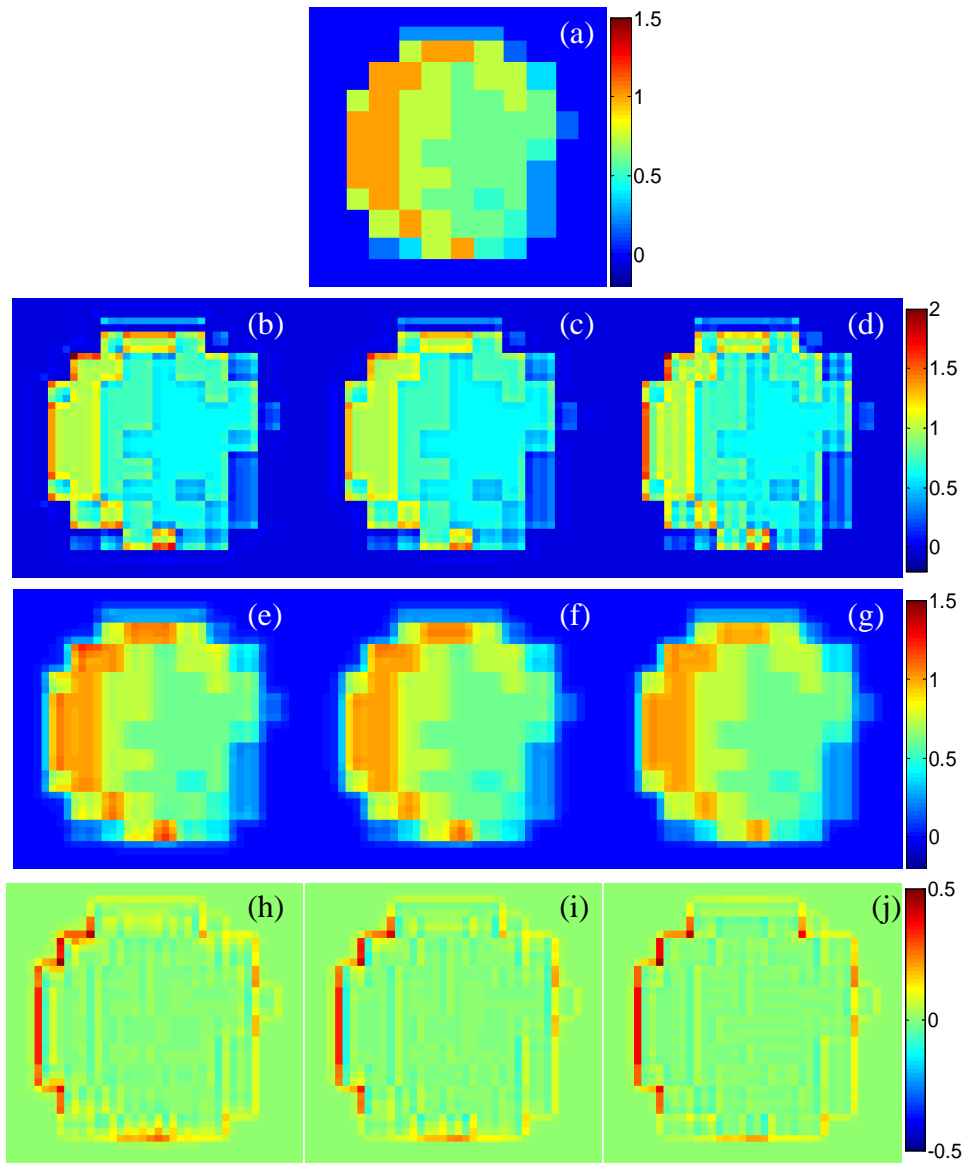
如图 4-7 所示，与其它两种方法相比，本章提出方法的卷积后注量图有较少热点和冷点。最明显的改进为减少了卷积后注量图和静态注量图在射野边缘的差值。我们用均方差来定量地衡量静态注量图和卷积后注量图之间的差别：

$$d = \left(\frac{1}{N} \sum \sum (\mathbf{F}_{ori} - \mathbf{F}_{con})^2 \right)^{1/2} \quad (4.26)$$

式中 N 为注量图内像素点的个数, \mathbf{F}_{ori} 为静态注量图, \mathbf{F}_{con} 为卷积后的注量图。图 4-7(b)-(d) 中的 3 种卷积后注量图与静态注量图的均方差依次为 3.50MU、2.35MU 和 0.30MU, 其中本章提出算法的卷积后注量图最接近静态注量图。

4.4.2.2 IMRT 注量图

我们对一个真实的 IMRT 进行反卷积, 其结果如图 4-8 所示。与静态注量图(图 4-8(a))相比, 鲁棒的注量图(图 4-8(b)-(d))在射野边缘处有突起。由于 x 轴方向的几何不确定性较大以及在 x 轴方向注量图的梯度比较大, 鲁棒注量图在 x 轴方向的突起尤为明显。图 4-8(e)-(g)和 (h)-(j)分别为反卷积核算法、改进的反卷积算法和新算法的卷积后注量分布及其与静态注量分布的差值。图 4-8(k)为卷积后注量图的剖面注量曲线。在不同反卷积算法所得的鲁棒注量图中, 新算法所得的鲁棒注量图不是最平滑的一个。但是, 其卷积后的注量图是最接近静态注量分布的一个, 尤其是在射野的边缘位置, 如图 4-8(h)-(j)所示。从图中我们可以观察到, 新算法明显地减小了卷积后注量图在射野边缘位置与静态注量图的差别。图 4-8(h)-(j)中 3 种卷积后注量分布与静态注量分布在射野内的均方差依次为 6.00MU、5.25MU 和 3.45MU。



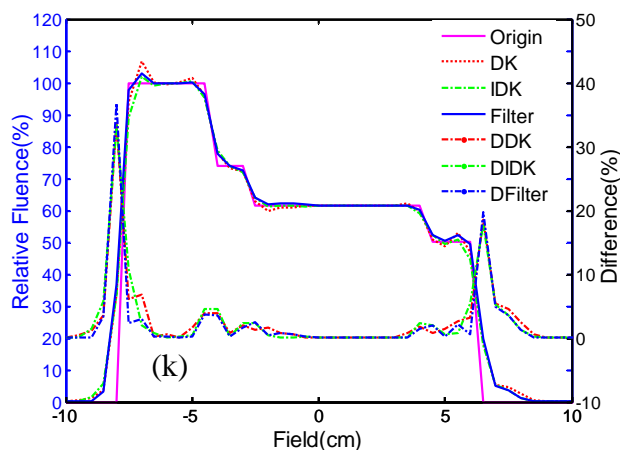
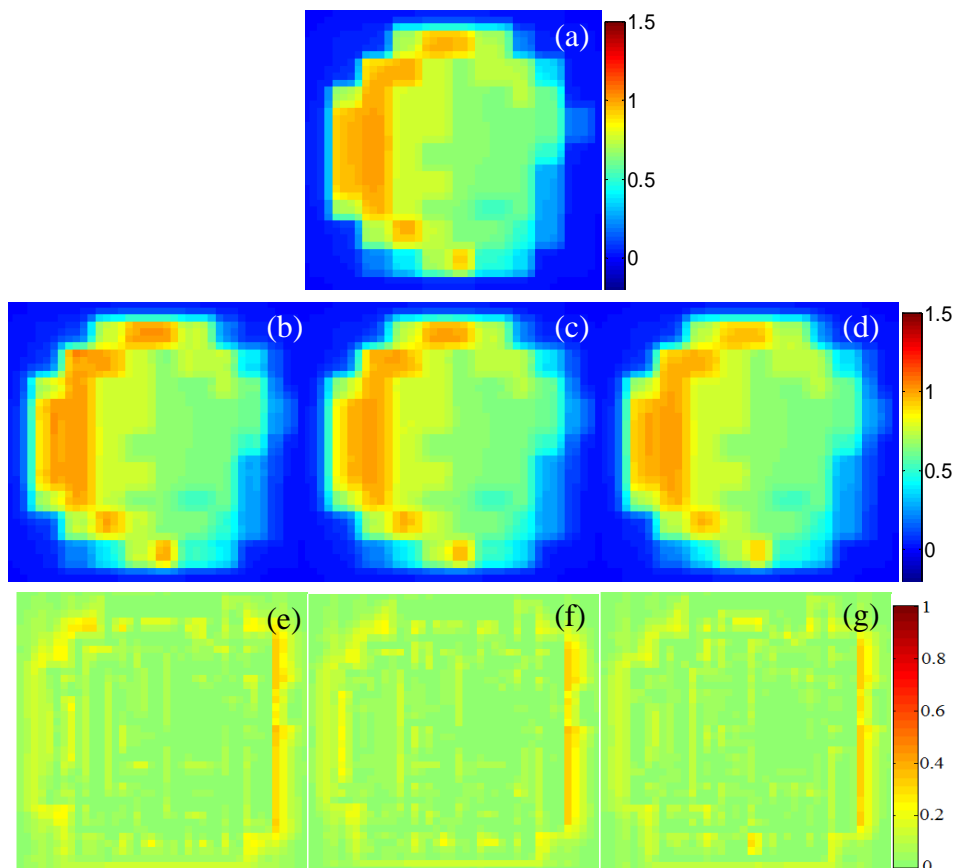


图 4-8 静态注量图、健壮注量图卷积后注量图及其剖面曲线和卷积后注量图与静态注量图的差值。(a)静态注量图；(b)反卷积核算法的健壮注量图；(c) 改进的反卷积核算法的健壮注量图；(d)新算法的健壮注量图；(e)-(g) 为健壮注量图(b)-(d)的卷积后注量图；(h)-(j)为卷积后注量图(b)-(d)与静态注量图的差值；(k) 卷积后注量图(b)-(d)的剖面曲线。

我们利用软件 CERR^[21] 计算图 4-8 中静态注量图 and 卷积后注量图在水体模内的剂量分布。在 6MV 多能射线照射的情况下，我们分别计算不同注量图在体模中距离表层 10cm 处的剂量分布，如图 4-9 所示。图 4-9(a)-(d)依次为静态注量图(图 4-8(a))和卷积后注量图(图 4-8(e)-(g))的剂量分布。我们进一步利用 gamma 分析算法来定量地分析不同卷积后注量图的剂量分布与静态注量图的剂量分布之间的差异。在使用 gamma 算法时，剂量差别参数和距离参数分别为 3%和 3mm。图 4-9(e)-(g)为卷积后注量图的剂量分布的 2D gamma 分析结果。图 4-9(h)为不同剂量分布的剖面曲线。从图中我们可以观察到，在 3 种剂量分布(图 4-9(e)-(g))中，利用新算法计算所得剂量分布最接近静态注量图的剂量分布，尤其是在射野的边缘位置。



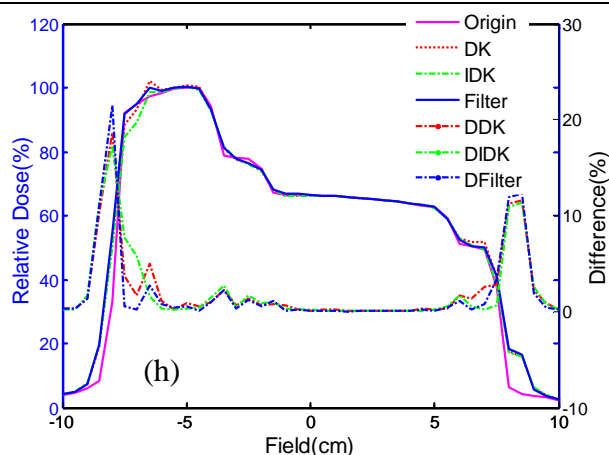


图 4-9 剂量分布及其剖面曲线和 γ 分析结果。(a)-(d)依次为静态注量图和图 4-8(e)-(g)卷积后注量图的剂量分布, (e)-(g)为将卷积后注量图的剂量分布(b)-(d)分别与静态注量图的剂量分布进行 γ 分析所得结果, (k) 剂量分布的剖面曲线。

4.4.2.3 前列腺肿瘤病例

我们在一例前列腺肿瘤患者的 CT 数据上验证本章提出算法的有效性。我们从 6 MV 多能射线在 5 个射野方向进行共面照射的 IMRT 计划中, 提取 90°射野方向的注量图。射线源到等中心点的距离为 100cm。图 4-10 为 CT 数据中的一层断层成像和提取出来的静态注量图。在前列腺肿瘤病例中, 我们考虑的正常组织为直肠壁和膀胱壁, 如 4-10(a)所示。我们分别计算了静态注量图、边缘扩展方法所得注量图和不同反卷积算法卷积后注量图的剂量分布。通过将静态注量图向各个方向外扩 5mm, 我们得到射野扩展方法的注量图。图 4-11 为不同注量图在体模内距离表层 17cm 处的剂量分布。从图中我们可以观察到, 使用边缘扩展法所得注量图对病人进行照射, CTV 区域之外的正常组织也接受到高剂量的照射, 如图 4-11(b)所示。使用反卷积核算法, 在靶区周围仍接受到高剂量的照射。相比边缘扩展的方法, 改进的反卷积核方法和本章提出的方法减少了正常组织接收到的剂量分布, 增加了靶区剂量的适形和调强程度。图 4-12 为 2D 剂量分布(图 4-11)的剂量-体积曲线, 其中射野扩展的方法被标识为“Margin”。相比其它 3 种方法, 新算法降低了靶区周围正常组织的剂量分布(图 4-12(a)和(b)), 保护了正常组织。对于 CTV, 使用新方法所得剂量分布最接近静态注量分布的剂量分布(图 4-12(c))。同样, 我们使用 γ 分析方法对使用不同方法所得剂量分布与静态注量图的剂量分布进行分析。如图 4-13 所示, 新算法明显地减少了卷积后注量图的剂量分布与静态注量图的剂量分布的差异。

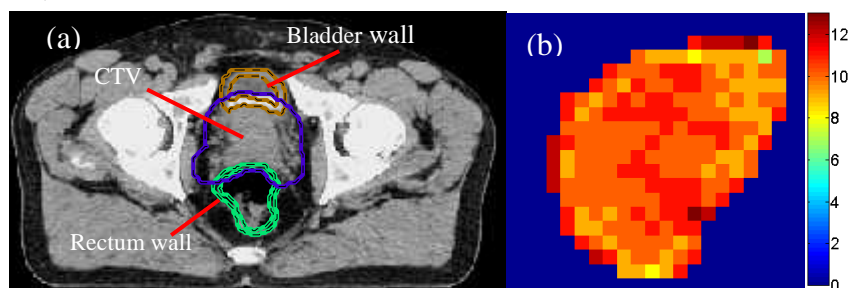


图 4.10 CT 数据和注量图。(a)CTV 和正常组织在 CT 数据内的相对位置, (b)从计划中提取的静态注量图。

我们使用 Engel 等人^[23]提出的方法将注量图分割成机器可执行的多叶准直器序列。在分割过程中, 我按注量图像素大小为 $0.5 \times 0.5\text{cm}^2$, 子野最小的机器跳数为 5MU (monitor unit)

对不同算法所得的鲁棒注量图进行分割。因此，分割后所得多叶准直器序列是机器可执行的。分别对边缘扩展算法所得注量图和反卷积核算法、改进的反卷积算法与新算法计算所得鲁棒注量图进行分割，所得子野个数分别为 10、17、14 和 16，所需总的机器跳数分别为 85MU，275MU，175MU 和 210MU。

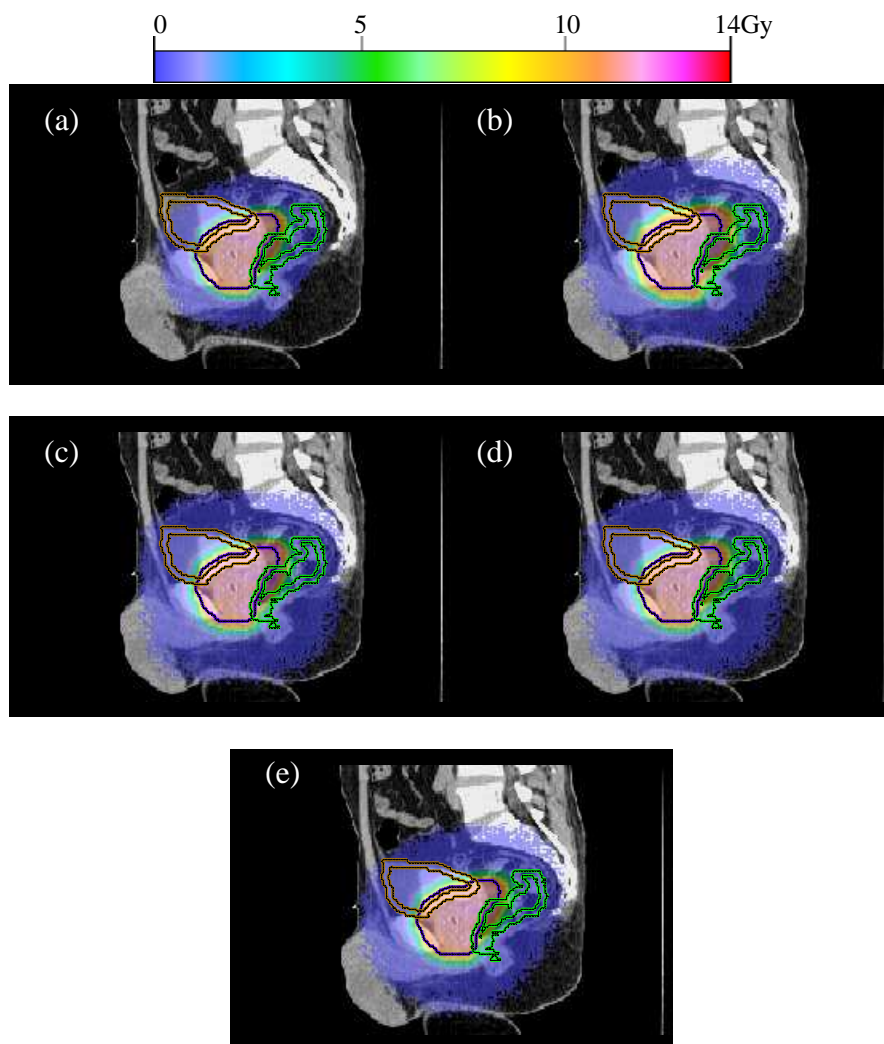
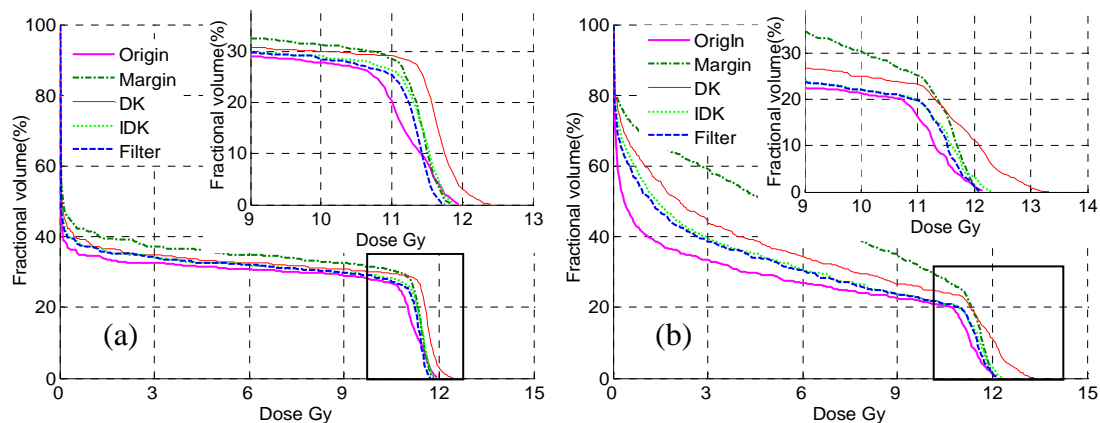


图 4-11 利用不同方法所得注量图的剂量分布。(a)静态注量图；(b)边缘扩展法；(c) 反卷积核算法；(d) 改进的反卷积核算法；(e)本章提出新算法。



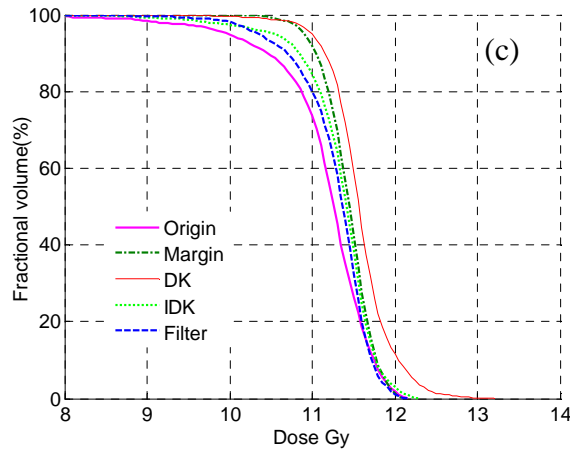


图 4-12 不同剂量分布下的剂量-体积曲线。(a)膀胱壁；(b)直肠壁；(c) CTV。

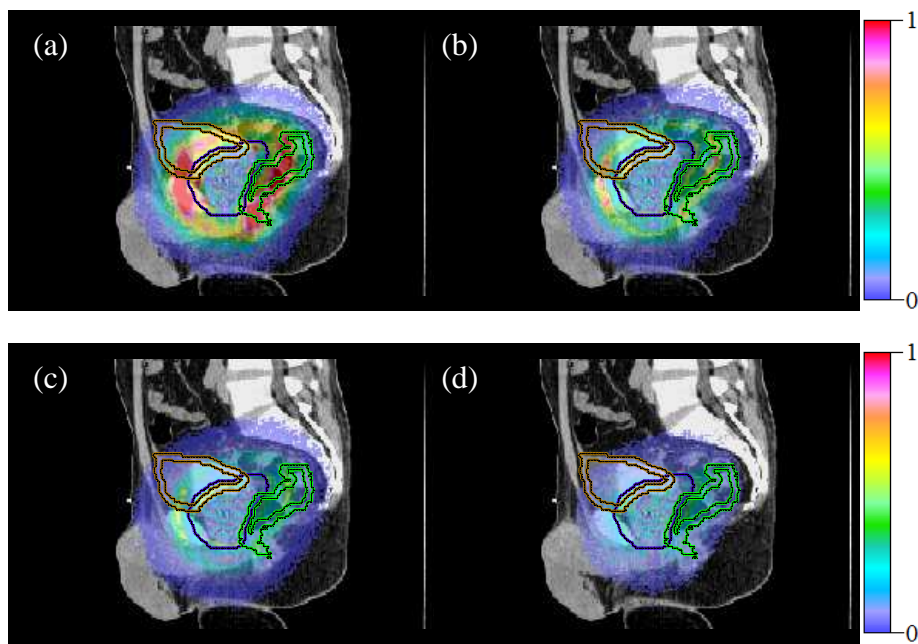


图 4-13 不同方法所得剂量分布与静态注量图的剂量分布的 gamma 分析结果。(a)边缘扩展法；(b)反卷积核法；(c)改进的反卷积核法；(d)本章提出的新方法。

4.4.3 讨论

本章的工作是在假设病人几何不确定性的概率密度函数是已知的基础上进行的。在此基础上，已有一些文献中报道了计算鲁棒注量图的方法^[4-8, 16]。Fan 等人^[16]以讨论了使用反卷积方法计算鲁棒注量图的优点和缺点。在文献中，Fan 等人^[16]表明序列反卷积的方法是适合于对 IMRT 注量图进行反卷积得到鲁棒的注量图，但是要求病人随机几何不确定性的运动幅度不能太大，概率密度函数的标准差一般要求在 2 到 4mm 之间。在本章中，我们对已有的序列反卷积算法进行分析，提出一种新的反卷积算法来求鲁棒注量图。

本章提出的算法由直接序列展开算法和滤波函数两部分组成。根据滤波器的不同性能，如通带性能、滚降因子(描述过渡带下降程度的一个因子)、相位响应和运行速度等，我们可以选择不同的滤波器，如贝塞尔滤波器、切比雪夫滤波器、椭圆滤波器等。例如，贝塞尔滤波器的过渡带比较窄，但是它的相位响应不是最好的。由于巴特沃斯滤波器平坦的通带以及较小的滚降因子，在本章中我们使用巴特沃斯滤波器进行滤波。本章主要目的是验证新算法的有效性。在将来的工作中，我们将比较使用不同滤波器情况下的反卷积结果。

卷积后的注量图是在考虑随机运动的情况下病人实际接收到的注量分布。我们的目的是

计算鲁棒注量图,使得卷积后注量图与病人静止不动时接收到的静态注量图相同,即尽可能地减少二者之间的差异。与反卷积核算法^[17]和改进的反卷积核算法^[16]相比,新算法明显地减小了卷积后注量图与静态注量图差异。在实际的 IMRT 数据处理中,关于这方面优点的报道是很少见的。但是,新方法增加了鲁棒注量图的复杂度。边缘扩展方法所得注量图和反卷积核算法、改进的反卷积核算法以及新算法计算所得鲁棒注量图的复杂度依次为: 83.62MU、268.05MU、179.26MU 和 206.59MU。

4.5 本章小结

在本章中,我们首先介绍了 IMRT 中的几何不确定性,然后提出了一种基于直接序列展开算法和频域滤波的混合反卷积方法来求解鲁棒注量图,并在两个 2D 注量图和一个前列腺肿瘤病例上验证了新算法的有效性。实验结果表明,利用巴特沃斯滤波器可以很好地滤除直接序列展开算法反卷积结果中的高频成份,减少卷积后注量图中的冷点和热点。在注量图内的平滑区域,新算法可以取得和其它反卷积算法相同的效果。在射野的边缘位置,新算法明显地减少了卷积后注量图与静态注量图的差异。新算法易于在临床实践使用,改进靶区剂量的均一性,提高放射治疗质量。

4.6 参考文献

- [1] International Commission on Radiation Units and Measurements. ICRU Report 62. Prescribing, recording, and reporting photon beam therapy (Supplement to ICRU Report 50). Bethesda, MD: ICRU, 1999.
- [2] van Herk M. Errors and margins in radiotherapy [J]. Semin Radiat Oncol, 2004, 14(1):52-64.
- [3] Yan D, Ziaja E, Jaffray D, et al. The use of adaptive radiation therapy to reduce setup error: a prospective clinical study [J]. Int J Radiat Oncol Biol Phys, 1998, 41(3):715-720.
- [4] Witte MG, van der Geer J, Schneider C, et al. IMRT optimization including random and systematic geometric errors based on the expectation of TCP and NTCP [J]. Med Phys, 2007, 34(9):3544-3555.
- [5] Craig T, Wong E, Bauman G, et al. Impact of geometric uncertainties on evaluation of treatment techniques for prostate cancer [J]. Int J Radiat Oncol Biol Phys, 2005, 62(2):426-436.
- [6] Li XA, Qi XS, Pitterle M, et al. Interfractional variations in patient setup and anatomic change assessed by daily computed tomography [J]. Int J Radiat Oncol Biol Phys, 2007, 68(2):581-591.
- [7] Santanam L, Esthappan J, Mutic S, et al. Estimation of setup uncertainty using planar and MVCT imaging for gynecologic malignancies [J]. Int J Radiat Oncol Biol Phys, 2008, 71(5):1511-1517.
- [8] Moore JA, Gordon JJ, Anscher MS, et al. Comparisons of treatment optimization directly incorporating random patient setup uncertainty with a margin-based approach [J]. Med Phys, 2009, 36(9):3880-3890.
- [9] van Herk M, Remeijer P, Lebesque JV. Inclusion of geometric uncertainties in treatment plan evaluation [J]. Int J Radiat Oncol Biol Phys, 2002, 52(5):1407-1422.
- [10] Lind BK, Källman P, Sundelin B, et al. Optimal radiation beam profiles considering uncertainties in beam patient alignment [J]. Acta Oncol, 1993, 32(3):331-342.
- [11] Löf J, Lind BK, Brahme A. Optimal radiation beam profiles considering the stochastic process of patient positioning in fractionated radiation therapy [J]. Inverse Problems, 1995, 11(6):1189-1209.
- [12] Baum C, Alber M, Birkner M, Nüsslin F. Robust treatment planning for intensity modulated radiotherapy of prostate cancer based on coverage probabilities [J]. Radiother Oncol, 2006, 78(1):27-35.

- [13] Chan TC, Bortfeld T, Tsitsiklis JN. A robust approach to IMRT optimization [J]. *Phys Med Biol*, 2006, 51(10):2567-2583.
- [14] Ólafsson A, Wright SJ. Efficient schemes for robust IMRT treatment planning [J]. *Phys Med Biol*, 2006, 51(21):5621-42.
- [15] Bortfeld T, Chan TCY, Trofimov A, et al. Robust management of motion uncertainty in intensitymodulated radiation therapy [J]. *Oper Res*, 2008, 56(6):1461-1473.
- [16] Fan Y, Nath R. Intensity modulation under geometrical uncertainty: a deconvolution approach to robust fluence [J]. *Phys Med Biol*, 2010, 55(14):4029-4045.
- [17] Ulmer W, Kaissl W. The inverse problem of a Gaussian convolution and its application to the finite size of the measurement chambers/detectors in photon and proton dosimetry [J]. *Phys Med Biol*, 2003, 48(6):707-727.
- [18] Ulmer W. Inverse problem of linear combinations of Gaussian convolution kernels (deconvolution) and some applications to proton/photon dosimetry and image processing [J]. *Inverse Problems*, 2010, 26(8):1-26.
- [19] Fang MT, Shei SS, Nagem RJ, et al. Convolution and deconvolution with Gaussian kernel [J]. *Nuovo Cimento B*, 1994, 109(1):83-92.
- [20] García-Vicente F, Delgado JM, Rodríguez C. Exact analytical solution of the convolution integral equation for a general profile fitting function and Gaussian detector kernel [J]. *Phys Med Biol*, 2000, 45(3):645-650.
- [21] Deasy JO, Blanco AI, Clark VH. CERR: a computational environment for radiotherapy research [J]. *Med Phys*, 2003, 30(5):979-985.
- [22] Low DA, Harms WB, Mutic S, et al. A technique for the quantitative evaluation of dose distributions [J]. *Med Phys*, 1998, 25(5):656-661.
- [23] Engel K. A new algorithm for optimal multileaf collimator field segmentation [J]. *Discrete Appl Math*, 2005, 152(2005):35-51.

第五章 自适应放射治疗中确定最佳重新规划时间及次数算法的研究

5.1 引言

随着外照射分次放射治疗技术的发展,为临床放射治疗增加了提供高度适形强辐射剂量的能力,使高剂量分布紧紧地覆盖肿瘤区域,如 3D 适形技术和调强放射治疗技术(IMRT)。临床研究以及模拟实验表明,利用高度适形的放射治疗技术可以减少肿瘤细胞的扩散和降低正常组织并发症发生的概率^[1]。随着功能成像技术使用的增加,对患者的保护中更包含了生物信息和功能信息,从而使得放射治疗计划变得更加复杂^[2]。高度适形的剂量分布,使得被照射患者体内的剂量梯度变得更加陡峭。从而增加了对病人在放射治疗时的摆位精度及多次治疗时的重复摆位精度的要求。因此,在治疗室内利用成像工具获取病人的解剖结构信息,并根据这些信息调整病人位置、射野形状及权重等其它参数的自适应放射治疗(adaptive radiotherapy, ART)方法,是完全实现调强放射治疗优点的关键技术。

近几十年来,医学成像技术在放射治疗质量保证方面发挥着越来越重要的作用。图像引导(image-guided radiation therapy, IGRT)技术根据病人的影像数据,在治疗室内对病人进行平移或旋转,使得病人位置与计划中的位置相吻合。但是,在治疗过程中,病人体内的解剖结构随放射治疗时间不断地反生变化,仅使用实时移位的方法已不能完全地消除由病人体内器官形变以及相对位置改变所引起的误差,尤其是对那些在治疗过程中运动非常显著的器官。这时需要根据实时获取的病人 3D 数据调整射野形状,如呼吸门技术和靶目标跟踪技术。然而,实时移位和改变射野形状不能完全校正由于分次照射引起的靶目标和正常组织内的剂量变化。这时需要根据剂量的变化对初始计划进行修正或重做计划。剂量引导(dose-guided)ART 技术主要有:在线 ART 和离线 ART。在线 ART 是根据当前体模结构相对于初始体模结构的变化,对初始计划进行实时修改,包括对射野形状及权重的调整。在线 ART 主要是对当前分次照射引起的随机误差进行校正。离线 ART 是通过重新规划来校正靶目标和正常组织内已照射剂量与理想剂量分布的偏差。离线 ART 主要是对前面分次照射中的系统误差进行校正。

离线 ART 技术被用来有效地校正由病人体内解剖结构的变化所引起的放射治疗误差。然而,离线 ART 所需的工作量特别大,包括进行 CT 扫描、勾画器官、计算累积剂量和重新制定计划。由于医疗资源有限,不可能为每个病人在每次分次放射治疗前做一次离线 ART 计划。然而,在临床可接受的范围内,我们可以适当地减少重新规划的次数而不影响放射治疗的质量。因此,有必要确定离线 ART 中的最佳重新规划的次数及相应时间。已有一些学者从不同层面描述了确定最佳一次重新规划的时间^[3-7],如根据体模内解剖结构的变化或剂量的变化。利用解剖结构变化进行确定的方法^[4,6],是根据病人体内解剖结构变化最大的时刻来确定最佳重新规划时间。而根据剂量变化确定最佳重新规划时间的方法^[3,5],是利用初始计划在分次照射时获取的 CT 数据上计算剂量分布,找到治疗质量最差的分次照射,并以此来决定最佳的重新规划时间。但是,新计划还被应用到其后所有的分次照射上,不能确定该新计划的整体输出是最优的。因此,这种方法不是非常合理。Wu 等人^[7]分别在不同的分次照射之前进行重新规划,并根据整体的剂量输出来确定最佳的重新规划时间。在本章中,我们采用类似的做法,根据整体治疗效果最优的组合确定最佳重新规划的时间及次数。Wu 等人^[7]的工只讨论了一次重新规划情况下的最佳重新规划时间,而在本章的工作中没有限制重新规划的次数。

本章首先介绍了 ART 的组成部分并给出计算累加剂量的方法。然后,以头颈部肿瘤为例,提出了一种确定最佳重新规划时间及次数的方法。最后,我们给出头颈部肿瘤自适应放射治疗中的最佳重新规划时间及次数。

5.2 自适应放射治疗技术

ART 技术的应用是以三个主要部分为基础^[8]:图像引导技术、剂量验证和计划调整,如图 5-1 所示。

5.2.1 图像引导

随着科学技术的发展,成像技术在放射治疗领域所发挥的作用越来越重要,从病情诊断、器官勾画、计划设计、放射治疗,到计划评估都要利用病人的解剖结构信息。在 ART 领域,图像引导技术主要提供了两方面的信息:1) 日常病人解剖结构数据,用来监视当前病人的解剖结构变化;2) 确定病人日常所得数据内解剖结构之间的差异。因此,图像引导在 ART 中的应用主要为:3D 图像获取和图像配准。日常解剖结构图像获取的方法可以分为 5 大类:螺旋 CT、千伏 CBCT(cone-beam CT)、兆伏 CBCT、兆伏 CT 系统和 Tomotherapy 系统^[3]。这些成像系统在不断的完善,利用这些成像系统可以精确确定肿瘤的位置,且软组织的分辨率可达 1-2%。

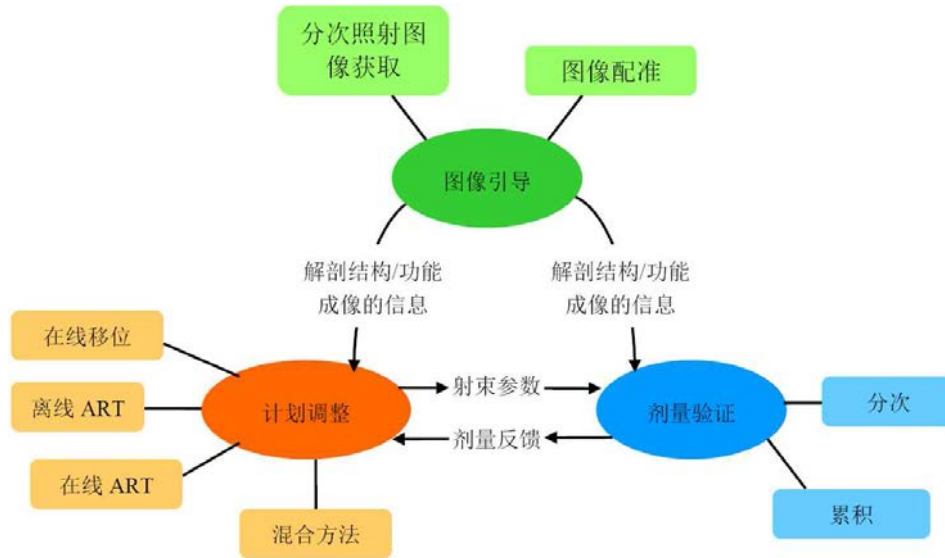


图5-1 ART技术的主要组成部分及相互关系^[8]。

图像配准技术主要是用来确定病人初始所得数据的解剖结构和分次照射时所得数据的解剖结构之间的差异。图像配准方法可以分为刚性配准和形变配准。刚性配准是通过 3D 移动,将分次照射时获取的图像对齐到初始图像,如病人初始所得 CT 数据。刚性配准可以被用来移动病人位置,以使靶目标对准。由于这种将靶目标对齐的方法速度快且鲁棒性比较好,因此在 IGRT 的临床实践中被广泛使用。但是,当靶目标或正常组织发生形变时,刚性配准不能提供足够多的形状和体积变化信息^[9]。因此,形变配准被用来确定两组数中体素到体素的映射关系。Yan 等人^[10]采用弹性物体的生物力学模型进行形变配准,并提出了一个计算放射治疗过程中各器官累积剂量的框架。Colsemon 等人^[11]提出了一种基于互信息的 CT 图像配准方法。Lu 等人^[12]利用基于自由范数的形变算法将千伏 CT 数据和兆伏 CT 数据进行配准。他们指出,在有很大的拓扑结构变化的情况下,基于密度的形变配准算法可能会失效。Kashani 等人^[13]在多器官体模内研究了不同的形变配准算法,结果表明,对于不同的算法,大的配准误差会出现在体模内不同的区域;对于同一种配准算法,不同的参数设置、输入数据等,同样会导致配准结果有大的变化。因此,在临床应用中,任何配准算法在使用前都需要验证其应用环境以及所用参数。

随着肿瘤放射治疗技术的发展,越来越多的生物和功能信息被用到病人的护理过程中。物理师和医生对功能成像方式获取的信息越来越感兴趣,如对比度增强的 MRI (magnetic resonance imaging)和 PET (positron emission tomography)。一些验证性的实验表明,使用功能成像引导的放射治疗可以进一步改善放射治疗质量^[14, 15]。在 ART 中使用功能成像还处于初级阶段,而且存在着许多技术难题和实验数据的限制。虽然在理论上功能成像提供的额外生物信息是非常有用的,但是在临床实践上使用这些技术之前,需要进行大量的系统分析来证明

功能成像带来的优点。

5.2.2 计划调整

计划调整是自适应放射治疗技术的重要组成部分之一，利用图像引导获取的图像信息来减少病人实际解剖结构和计划中解剖结构之间的位置偏移，从而增加了靶区剂量覆盖和尽可能地保护正常组织不受照射。学者们已提出了一些算法或方案进行计划调整^[15-27]，部分方法已开始在临床上使用。这些方法可以被分为 4 类。

5.2.2.1 在线移位

在线移位是实现 ART 的第一阶段，并在临床上广泛的应用。临床上实现在线移位需要 IGRT 技术的支持。在每次分次照射时，根据获取的病人影像数据对病人的位置进行在线调整。在线移位主要用来校正当前病人的靶区相对于计划中靶区的位置偏移。不同类型的标记点被用来对齐靶区的位置。在治疗过程中部分靶区的位置相对于骨骼的位置是不变的，如颅内肿瘤。因此，在大多数的商用治疗软件中都提供在线骨骼结构勾画功能，用来在线移动病人的位置。但是，还有一些肿瘤相对于骨骼位置的变化比较明显，是不适合使用在线移动的方法，如肺和腹。为了精确地确定靶区的位置，注册标记物(fiducial marker)被植入病人体内的软组织，用来代替靶目标的位置。Chung 等人提出了利用注册标记物的方法对前列腺肿瘤患者靶区进行在线移位和对准的方法，该方法在临床实践上被广泛使用^[16]。Jaffray 等人利用 CBCT 获取的病人数据对软组织位置进行对齐，从而减少靶区位置的不确定性^[17]。Wu 等人利用病人的几何信息以及体内剂量分布信息对在线移位的方法进行了评价，并证明了利用在线移位方法缩小由 CTV 到 PTV 扩展部分的有效性^[18]。在线移位方法的效果比较好而且在临床上广泛的使用，但是这种方法不能完全地消除由分次照射间靶目标或正常组织形变引起的靶区剂量不足，例如前列腺和头颈部肿瘤的放射治疗。对于靶区或正常组织的形变引起的剂量误差，需要重新规划的方法来调整计划。

5.2.2.2 离线 ART

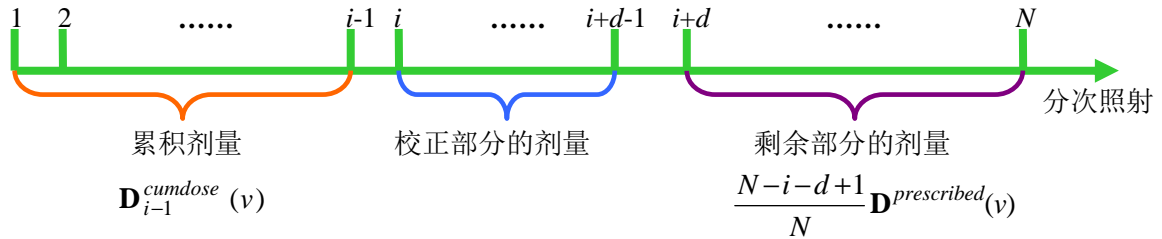


图5-2 离线ART技术的原理。

Yan 等人首先在 ART 中提出了离线重新规划的概念^[19]。放射治疗进行一段时间后，初始的计划可能不再满足临床需求，这时需要重新规划，并在随后分次照射中校正之前照射引起的剂量不足，如图 5-2 所示。按照理想的放射治疗计划，经过 N 次分次照射完成放射治疗，体模内的总剂量分布为 $D^{prescribed}(v)$ ，即每次分次照射在体模内产生的理想分次剂量分布为

$\frac{1}{N} D^{prescribed}(v)$ ，其中 v 为体模内的体素。但是，由于病人体内器官在放射过程中的移位、形变以及其它因素，每次分次照射在体模内的剂量分布并不能达到理想分次剂量分布。假设从第 i 次分次照射时开始考虑校正前面 $i-1$ 次分次照射中引起的剂量误差。前面 $i-1$ 次分次照射在体模内的实际剂量分布 $D_{i-1}^{cumdose}(v)$ ，而理想 $i-1$ 次分次照射在体模内的实际剂量分布为

$\frac{i-1}{N} D^{prescribed}(v)$ 。因此，前面 $i-1$ 次分次照射中引起的剂量误差为 $\frac{i-1}{N} D^{prescribed}(v) - D_{i-1}^{cumdose}(v)$ 。

知道剂量误差之后，就可以重新设计计划，在从 i 次分次照射开始的 d 次分次照射内，校正

前面分次照射引起的剂量误差, d 的取值范围可以从 1 到 $N-i+1$ 。利用 d 次分次照射校正前面分次照射引起的剂量误差是基于已知这 d 次的 CT 数据的假设, 从而才能完全校正这些误差。已有两种方法获取这 d 次分次照射的 CT 数据: 1) 认为与 $i-1$ 分次照射后得到的 CT 数据相同, 既在这段时间内, 体模内的器官没有发生形变; 2) 根据前面 $i-1$ 次分次照射的 CT 数据进行估算得到。Wu 等人^[20]和 de la Zerda 等人^[21]就是按照这种方法分别提出了离线 ART 的方法。Wu 等人^[20]首先提出了用物理准则表示剂量误差。de la Zerda 等人^[21]认为离线 ART 是一个反馈系统, 并根据校正剂量误差所用不同的分次照射次数 d 分别进行讨论。但是, 这种方法是以前 $i-1$ 次分次照射的剂量误差为基础提出来的, 不能直接扩展到关于全局剂量分布的生物准则上。

如图 5-2 所示, 体模内的剂量分布可以由 3 部分组成: 已照射剂量分布, 校正剂量误差的分次照射中的剂量分布和剩余分次照射的理想剂量分布, 其公式为:

$$\mathbf{D}(v) = \mathbf{D}_{i-1}^{cumdose}(v) + \mathbf{D}_d^{correction}(v) + \frac{N-i-d+1}{N} \mathbf{D}^{prescribed}(v) \quad (5.1)$$

其中 $\mathbf{D}_d^{correction}(v)$ 为在离线重新规划后的 d 次分次照射中的总剂量分布:

$$\mathbf{D}_d^{correction}(v) = \mathbf{D}_i^{correction}(v, \psi_i) + \dots + \mathbf{D}_{i+d-1}^{correction}(v, \psi_{i+d-1}) \quad (5.2)$$

其中 ψ_i 为 i 次分次照射时使用的照射计划, $\mathbf{D}_i^{correction}(v, \psi_i)$ 为该计划在体模内的剂量分布。离线 ART 就是要确定这 d 次分次照射的照射计划, 从 ψ_i 到 ψ_{i+d-1} 。如果假设在 d 次分次照射中体模内解剖结构没有反生变化, 则这 d 个分次照射的照射计划相同:

$$\mathbf{D}_d^{correction}(v) = d\mathbf{D}_i^{correction}(v, \psi_i) \quad (5.3)$$

在知道总剂量分布的情况下, 可以利用物理准则 F 或生物准则 G 来评价前 $i-1$ 次分次照射引起的误差, 并将其反馈给重新规划的目标函数。对于物理准则 F , 直接按 $F(\mathbf{D}^{prescribed}(v) - \mathbf{D}(v))$ 进行评价, 所得结果与 Wu 等人^[20]和 de la Zerda 等人^[21]提出方法的结果相同。由于生物准则是关于整体剂量分布的放射生物效应, 因此对于生物准则 G , 按 $G(\mathbf{D}^{prescribed}(v)) - G(\mathbf{D}(v))$ 的方式评估前 $i-1$ 次分次照射引起的误差。由于离线 ART 不能够在实时得到的 CT 数据上进行重新规划, 因此近似地估算校正剂量误差所使用的分次照射 CT 数据显得尤为重要^[22-24]。

5.2.2.3 在线 ART

在线 ART 是在分次放射治疗前获取的体模数据上进行计划修改, 校正当前分次放射治疗时由于体模内解剖结构的变化而引起的放射治疗误差。使用在线 ART 技术时, 病人在治疗床上等待物理师实时修改计划并用修改后的计划进行当前的分次放射治疗。因此, 在线计划修改需要在较短的时间内完成。由于重新做计划(两步法或直接子野优化法)所需的计算时间过长, 一些学者提出了不同的对初始计划进行修改的在线 ART 方法^[25-28]。Mohan 等人^[25]提出了根据靶区及周围正常组织在每个射野方向的形变改变多叶准直器位置的方法。首先根据分次照射 CT 数据和初始 CT 数据中靶区及周围正常组织在射野方向上的差异, 计算每个射野方向上的形变矩阵; 然后利用形变矩阵对初始计划中每个射野形状进行改变, 得到当前分次照射射野的形状。Court 等人^[26]也提出了一种可行的、逐片调整准直器叶片位置方法。但是这些方法只考虑了体模内器官在射野方向的 2D 形变, 而没有考虑各器官在体模内的 3D 形变。没有使用 3D 配准技术, 很难综合地考虑正常组织和靶区重合区域内的剂量分布。Feng 等人^[27]直接计算 3D 形变矩阵, 并根据该形变矩阵对各射野方向的子野形状进行修改。这种方法的

优点在于 3D 形变图像配准技术被用来获取体模内各器官的 3D 形变信息,保证了所有子野形状改变的一致性。前面所述方法都是针对各器官在射野的剖面方向对射野形状进行修正,不能对射野深度方向的形变进行校正。射野深度方向形变的校正主要是通过修改射野权重的方法来实现。Ahunbay 等人^[28]提出了一种对子野形状和权重同时修改的方法。该方法分 3 步来实现: 1) 在分次照射前得到的 CT 数据上迅速地勾画靶区和周围的正常组织,并计算形变矩阵; 2) 在线子野形状形变(segment aperture morphing, SAM),利用形变矩阵校正初始计划子野形状来确定当前分次照射的准直器叶片位置; 3) 子野权重优化(segment weight optimization, SWO)。

5.2.2.4 混合算法

在线移位的 ART 方法由于其操作简单、快速,在临床上被广泛应用。相对于在线移位的方法,离线 ART 方法和在线 ART 方法校正了由分次照射间体模内解剖结构变化而引起的放射治疗误差,提高放射质量。但是这两种方法也大大地增加了物理师和医生的工作量,限制了其在临床实践中的使用。在刚性靶目标变化或器官形变较小的情况下,在线移位仍然是一种非常有效的校正方法。在靶区形变较大的情况下,需要在当前分次照射得到的体模上进行重新规划来提高靶区剂量的覆盖以及将危及器官从高剂量区域内分离出来。因此,将在线移位方法和离线/在线 ART 方法组合成混合方法,可以在尽可能提高计划质量的同时减少重新规划所需冗长的时间^[29]。

5.2.3 剂量验证

ART 的过程可以被理解为一个闭循环系统,其中剂量验证作为系统的反馈。ART 中的剂量验证可以分为分次剂量验证和累积剂量验证两类。分次剂量验证主要是针对当前分次照射在体模内的剂量分布,并通过体模内的剂量分布来评价该分次照射计划的质量,用以判断当前计划是否可行。如果当前计划可行,则执行该计划;否则,将分次照射的剂量信息反馈回系统,并根据反馈信息调整初始计划。常用的方法是将当前分次照射的剂量-体积曲线与初始计划的剂量-体积曲线进行比较。分次剂量主要是被用来反馈系统的随机误差。累积剂量校正主要是针对当前分次照射之前的所有分次照射的累积剂量分布进行评价。由于在放射治疗过程中体模内的解剖结构在不断地发生变化,初始体模内的体素位置在随时间进行变化。通过剂量累积方法可以计算得到初始体模内每个体素在实际治疗过程中接收到的剂量。同样,累积剂量验证可以提供两方面的信息: 1) 是否需要在当前分次照射前进行重新规划; 2) 如果需要重新规划,将累积剂量和理想剂量的差值反馈并根据反馈信息调整计划。累积剂量主要是被用来反馈系统误差。

5.3 累加剂量

本章工作的主要内容是确定在离线 ART 过程中的最佳重新规划时间及次数。离线 ART 主要是根据系统反馈的累积剂量与离线剂量差来校正分次放射治疗过程中的系统误差。同样,我们使用累积剂量来评价计划的质量。剂量累加主要分三步来实现: 1) 将分次照射时得到的 CT 数据与初始 CT 数据进行配准并得到形变矩阵; 2) 利用得到的形变矩阵对分次照射的剂量分布进行校正; 3) 对校正后的剂量进行累加。

5.3.1 CT 数据配准

放射治疗过程中,由于体模内部的辐射效应引起了靶目标和正常组织的位置、形状及体积的显著改变。在医学领域,图像配准技术通常被用来监视肿瘤的生长情况或放射治疗的情况。配准对齐 CT 扫描数据,即通过变化将一组 CT 数据中的体素映射到达另一组 CT 数据的同源体素。在本章中,CT 数据配准分两步来实现: 1) 将分次照射获取的 CT 数据进行移位,使得处理后的数据与初始 CT 数据在同一坐标系统下,得到空间映射参数; 2) 将处理后的分次照射获取的 CT 数据与初始 CT 数据进行配准,得到形变矩阵。刚性配准技术被用来获取两

组 CT 数据之间的空间映射参数。非刚性配准算法, symmetric forces Demons 算法^[30], 被用来计算两组 CT 数据之间的形变矩阵。

对于给定的参考图像 **I** 和浮动图像 **J**, 图像配准的目标就是找到每一个像素点的最佳几何变换从而使两幅图像达到最佳匹配。基于灰度的配准方法, 利用图像的灰度信息, 采用某种搜索方法寻找形变模型的参数值, 使得建立在两幅图像之上的相似性度量值最大或最小。**s** 为无参数的几何变换, 则 **s(p)** 表示把浮动图像中的某一点 **p** 映射到参考图像空间中某一点。

一般用稠密的形变场 **s(p)** 来描述无参数的几何形变 **s**, 则 $\mathbf{s}: \mathbf{p} \rightarrow \mathbf{p} + \mathbf{s}(\mathbf{p})$ 。求解的基于图像灰度信息的图像配准问题通常可以表示为^[31]:

$$E(\mathbf{I}, \mathbf{J}, \mathbf{c}, \mathbf{s}) = \frac{1}{\sigma_i^2} \text{Sim}(\mathbf{I}, \mathbf{J}, \mathbf{c}) + \frac{1}{\sigma_x^2} \text{dist}(\mathbf{c}, \mathbf{s})^2 + \frac{1}{\sigma_T^2} \text{Reg}(\mathbf{s}) \quad (5.4)$$

其中 $\text{Sim}(\mathbf{I}, \mathbf{J}, \mathbf{c})$ 为相似性测度(similarity measure), 定量地衡量了两幅图像的匹配效果, 同时也衡量了当前变换的优劣, 该测度是图像配准过程中非常重要的一部分; σ_i 表示图像局部噪声的大小; $\text{Reg}(\mathbf{s})$ 为规则化项, 用来限制几何形变, 以产生稳定的解; σ_T 表示规则化的大小; **c** 为中间变量, 表示尚未规则化的空间几何变换, 从而使 Demons 算法被看成是对某一适应准则的优化问题; $\text{dist}(\mathbf{s}, \mathbf{c})$ 表示规则化前后两几何形变之间的距离; σ_x 表示 **s** 与 **c** 之间的不确定度。为了保证形变的可逆性并能够获得真正的可逆变换, symmetric forces Demons 算法^[30]在对数域内求解优化问题, 也就是说使用一个稳定的速度场 **v**, $\mathbf{s} = \exp(\mathbf{v})$ 。为了保证目标函数在对数域内的一致性和简化目标函数, 部分子目标函数被变换为: $\text{dist}(\mathbf{s}, \mathbf{c}) = \|\log(\mathbf{s}^{-1} \circ \mathbf{c})\|$ 和 $\text{Reg}(\mathbf{s}) = \|\nabla \log(\mathbf{s})\|^2$ 。在优化过程中, 通过求解光流方程:

$$E_{\text{diffeo}}^{\text{corr}}(\mathbf{I}, \mathbf{J}, \mathbf{v}, \mathbf{u}) = \text{Sim}(\mathbf{I}, \mathbf{J}, \exp(\mathbf{v}) \circ \exp(\mathbf{u})) + \|\mathbf{u}\|^2 \quad (5.5)$$

得到空间变化向量 **u**, 并根据 $\mathbf{v} = \mathbf{v} + \mathbf{u}$ 来更新 **v**。

通过计算在初始 CT 数据和配准后的分次照射获取的 CT 数据中靶区和周围正常组织的 Dice 值来验证 Demons 形变配准算法的精度。Dice 值的计算公式为:

$$\text{Dice} = \frac{2\mathbf{A}_{\text{organ}} \cap \mathbf{B}_{\text{organ}}}{|\mathbf{A}_{\text{organ}}| + |\mathbf{B}_{\text{organ}}|} \quad (5.6)$$

其中 $\mathbf{A}_{\text{organ}}$ 为器官在初始 CT 数据内的位置; $\mathbf{B}_{\text{organ}}$ 为该器官在配准后的分次照射获取的 CT 数据内的位置。

5.3.2 分次照射剂量分布校正

如上文所述, 使用图像配准的方法获得了两组 CT 数据之间的空间映射参数值和形变矩阵 **s**。对分次剂量分布的校正也是分两步来完成: 首先利用空间映射参数改变整体分次照射剂量分布的坐标位置, 保证处理后的分次照射剂量分布和初始剂量分布在同一坐标系统; 然后利用形变矩阵对处理后分次照射剂量分布进行变换, 使得校正后的分次照射剂量分布与初始剂量分布内的体素位置相互对齐。形变矩阵 **s** 是一个三维矩阵, 矩阵内的每个体素为一个三维向量, 表示处理后分次照射获取 CT 数据内的体素 **p_{in}** 相对于初始 CT 数据内的同源体素 **p_{out}** 的位置偏移:

$$\mathbf{p}_{\text{in}} = \mathbf{p}_{\text{out}} + \mathbf{s} \quad (5.7)$$

因此, 校正后的分次照射剂量分布 $\mathbf{D}_{T\text{-Fraction}}$ 可按下式获得:

$$\mathbf{D}_{T\text{-Fraction}}(\mathbf{p}_{\text{out}}) = \mathbf{D}_{D\text{-Fraction}}(\mathbf{p}_{\text{in}} - \mathbf{s}) \quad (5.8)$$

式中 $\mathbf{D}_{D-Fraction}$ 为处理后的分次照射剂量矩阵。在通常情况下，输出数据内的体素位置经过变换后不再为整数，这时需要利用插值算法近似地计算非整数位置处体素的剂量值，如 B 样条插值等。

5.3.3 剂量累加

依次经过以上两次对分次照射剂量分布进行处理之后，所有校正后的分次照射剂量分布都与初始剂量分布对齐。因此，将所有校正后的分次照射剂量分布相加就得到体模内每个体素的真实累积剂量。在分次照射后监视剂量分布情况所需的工作量比较大，包括获取 CT 数据、器官勾画、图像配准等。为了减少工作量，在连续多次分次照射后监测剂量分布的情况，并假设在连续两次监测范围内病人的解剖结构没有发生变化，即病人体内的剂量分布没有发生变化。剂量累加的公式写为：

$$\mathbf{D}_{cum} = \sum_{i=1}^N g_i \mathbf{D}_{T-Fraction}^i \quad (5.9)$$

式中 \mathbf{D}_{cum} 为累加剂量； N 为治疗过程中监测剂量分布的次数； $\mathbf{D}_{T-Fraction}^i$ 为第 i 次剂量监测时的校正后的分次照射剂量分布； g_i 表示从第 i 次剂量监测到第 $i+1$ 次剂量监测之间的分次照射次数。

5.4 头颈部肿瘤自适应放射治疗中的最佳重新规划时间及次数

5.4.1 实验设计

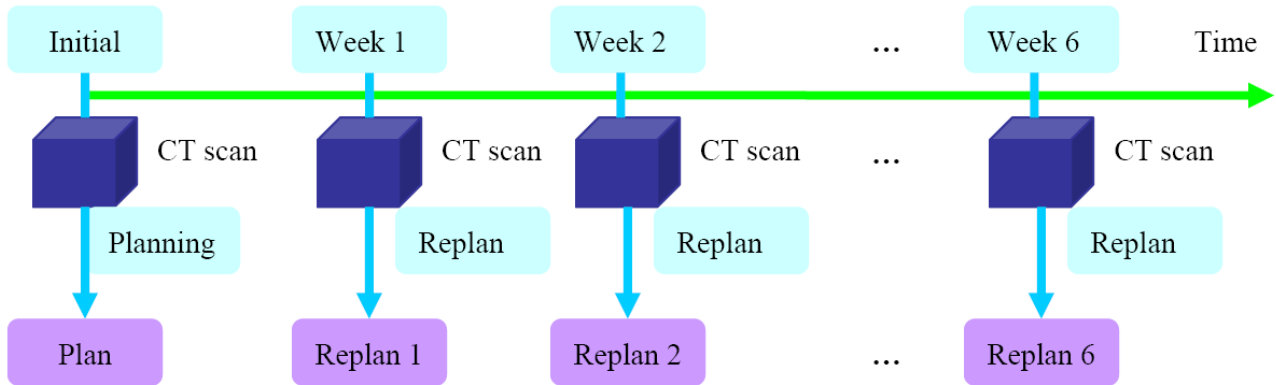


图 5-3 所获取的头颈部肿瘤病人数据。

在本章中，我们以头颈部肿瘤放射治疗为例，研究自适应放射治疗中的最佳重新规划时间及次数。自适应放射治疗技术是用来减少由于体模内解剖结构的形变而引起的剂量误差，其目的是在增加靶区剂量覆盖率的同时尽可能地将周围的正常组织从高剂量区域中分离出来。为了减少重新规划的工作量，在每周放射治疗结束后对病人进行重新规划，即每周重新规划一次。患者在初始计划完成 1 个月之后才开始接受放射治疗。在这一个月之内，由于肿瘤的生长，病人的解剖结构可能发生了大的变化。因此，在患者接收放射治疗之前有必要进行重新规划。剩余的重新规划都是在每周病人放射治疗结束后立即进行。我们所获取的头颈部肿瘤病人数据如图 5-3 所示。头颈部肿瘤患者的放射治疗一般在 5 到 7 周内完成。为了获取最佳的重新规划时间及次数，我们计算在不同的重新规划时间及次数下的累积剂量分布。例如，在 5 周的放射治疗过程中，不同重新规划时间及次数的所有组合形式有 31 种。其中，在不同时间进行重新规划 1 到 5 次的组合形式分别有 5、10、10、5 和 1 种。每种组合的具体重新规划时间如表 5.1 所示。例如，组合“Replan 2 4 5”表示该组合由 3 次重新规划组成，重新规划的时间分别为第 1、3 和 4 周分次放射治疗之后，并分别在第 2、4 和 5 周开始使用

新计划。需要注意的是重新做好的计划只能在之后的分次照射中使用，直到下次重新规划。

在已知每种组合具体重新规划时间的情况下，利用式 5.9 计算累积剂量。在肿瘤的放射治疗过程中，我们在每周设置一个剂量监测点，即在前一周放射治疗结束后，获取 CT 数据。如果在当前周使用重新规划后的计划，那么当前周内的分次照射剂量分布与新计划的剂量分布相同；否则，继续使用前一周所用的计划，这时，当前周内的分次照射剂量分布为前一周所用的计划在前一周放射治疗结束后获取 CT 数据上的剂量分布。不同组合的累积剂量分布是不同的，部分组合偏向于保护靶区剂量覆盖，而另一些组合则更有利于保护周围的正常组织。因此，我们设计了一个目标函数来综合地评价每种组合的累积剂量分布。总的目标函数由评价各器官的子目标函数及其权重乘积的组成：

$$F = \sum_{i=1}^{N^{organ}} w_i f_i(\mathbf{D}_{cum}^i) / \sum_{i=1}^{N^{organ}} w_i \quad (5.10)$$

其中， N^{organ} 表示所有要考虑器官的个数； \mathbf{D}_{cum}^i 表示第 i 个器官内的累积剂量分布； f_i 表示评价第 i 个器官内剂量分布的子目标函数，由物理准则或生物准则构成； w_i 为对应器官的权重系数。在本章中，我们只考虑自适应放射治疗方法对腮腺的保护作用，并使用平均剂量来评价腮腺内的剂量分布。平均剂量公式为：

$$f_{mean}(\mathbf{D}) = \frac{1}{N} \sum_{i=1}^N \mathbf{D}_i, \quad (5.11)$$

式中 N 为该器官内所有的体素个数， \mathbf{D}_i 为每个体素内的剂量值。因此，总的目标函数为两平均剂量函数的加权和，分别表示同侧(ipsilateral, IL)腮腺和异侧(contralateral, CL)腮腺。用集合 C 表示所有不同时间进行重新规划的组合(如表 5.1 所示)，最优的组合为所有组合中目标函数值最小的组合：

$$\min_{\beta \in C} (F_{\beta}) \quad (5.12)$$

最佳的重新规划时间及次数就是最优组合所使用的重新规划次数及所对应的重新规划时间。

表 5.1 在不同重新规划次数下不同重新规划时间的组合。

	一次	两次	三次	四次	五次
1	Replan 1	Replan 1 2	Replan 1 2 3	Replan 1 2 3 4	Replan 1 2 3 4 5
2	Replan 2	Replan 1 3	Replan 1 2 4	Replan 1 2 3 5	
3	Replan 3	Replan 1 4	Replan 1 2 5	Replan 1 2 4 5	
4	Replan 4	Replan 1 5	Replan 1 3 4	Replan 1 3 4 5	
5	Replan 5	Replan 2 3	Replan 1 3 5	Replan 2 3 4 5	
6		Replan 2 4	Replan 1 4 5		
7		Replan 2 5	Replan 2 3 4		
8		Replan 3 4	Replan 2 3 5		
9		Replan 3 5	Replan 2 4 5		
10		Replan 4 5	Replan 3 4 5		

本章的研究目的是确定自适应放射治疗过程中最佳的重新规划时间及次数。我们在从法国 Centre Eugène-Marquis 医院获取的 11 例头颈部肿瘤患者数据上进行了研究。重新规划在每周放射治疗结束后获取的 CT 数据上进行重新设计。初始计划和重新规划都是在 PINNACLE(Philips Medical Systems, Andover, MA)系统上实现的。在开源代码 ITK (Insight Segmentation and Registration Toolkit)和 VTK (Visualization Toolkit)进行 CT 数据配准、分次照

射剂量校正以及剂量累加。在放射治疗过程中,同一病人在初始 CT 数据及每周放射结束后获取 CT 数据上的器官轮廓由同一个医生进行勾画。其中,正常组织主要考虑同侧腮腺和异侧腮腺;靶区包括 PTV56Gy、PTV63Gy 和 PTV70Gy 三部分,即 3 个靶区的处方剂量分别为 56Gy、63Gy 和 70Gy。平均剂量和 NTCP 准则(式 3.18)被用来评价正常组织内的累积剂量分布。我们使用 Dijkema 等人^[32]在文献中给出的参数计算腮腺的 NTCP 值,其中 $n=1$, $m=0.4$, $TD_{50}=39.9\text{Gy}$ 。腮腺在不同 NTCP 模型参数下的剂量效应曲线如图 5-4 所示,其中 Burman 等人给出的参数来自文献[33]。

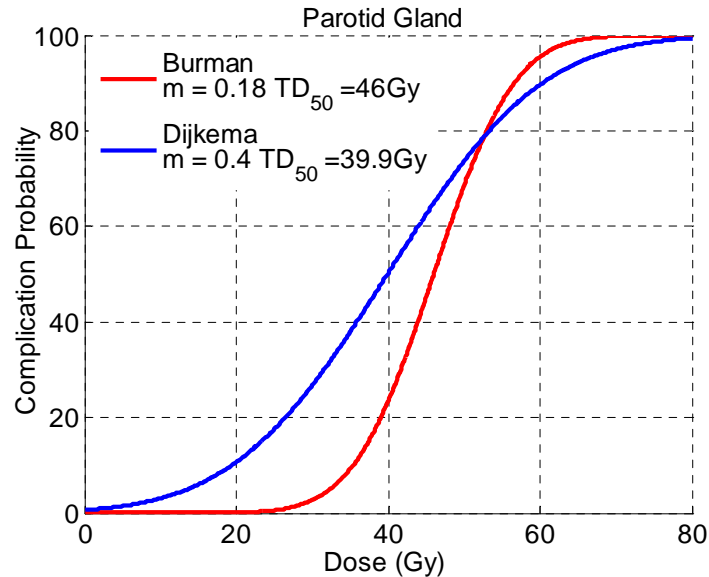


图 5-4 腮腺在不同 NTCP 模型参数下的剂量效应曲线。

5.4.2 实验结果

5.4.2.1 器官体积变化

在肿瘤放射治疗过程中,由于体模内的放射效应,部分器官的解剖结构发生了显著的变化,其中包括病人体重减少或增加、肿瘤收缩或扩张、淋巴结收缩或扩张、器官移位或形变等。靶区和周围正常组织的解剖结构变化是进行重新规划的主要原因。在本章中,我们记录了腮腺和 PTV 区域的体积变化,如图 5-5 所示。从图中我们可以观察到,腮腺的平均体积在随着放射治疗的进行而不断缩小;靶区 PTV56Gy 和 PTV63Gy 的平均体没有显著变化,PTV70Gy 的平均体积在放射治疗 2 周后开始缩小。

5.4.2.2 每周重新规划的剂量分布

重新规划的目的是在保证靶区剂量覆盖的情况下,尽可能的将正常组织从高剂量区域内分离出来。图 5-6 为 11 例头颈部肿瘤患者分别在初始剂量分布(Initial dose)、初始计划在每周获取 CT 数据上的剂量分布(Planned dose)和每周重新规划的剂量分布(Replanned dose)中两个腮腺组织的平均剂量值。从图中可以观察到,相对于初始计划在每周获取 CT 数据上的剂量分布,大部分每周重新规划的剂量分布减少了腮腺内的剂量分布。而剩余部分的重新规划,为了保证靶区剂量的分布,增加了腮腺组织内的平均剂量。每周的重新规划,减少了正常组织上的平均剂量。而正常组织上平均剂量的减少与体模内各器官的体积以及两器官之间的距离有密切的关系。每周重新规划剂量分布的平均剂量与初始剂量分布的平均剂量的差值和正常组织中心到各靶区中心距离的关系如图 5-7 所示。从图中可以观察到,随着正常组织中心到靶区中心距离的增加,重新规划后正常组织内的平均剂量与初始剂量分布的平均剂量的差值越小(如图 5-7 (b)和(c) 所示),重新规划对正常组织的保护作用越明显。

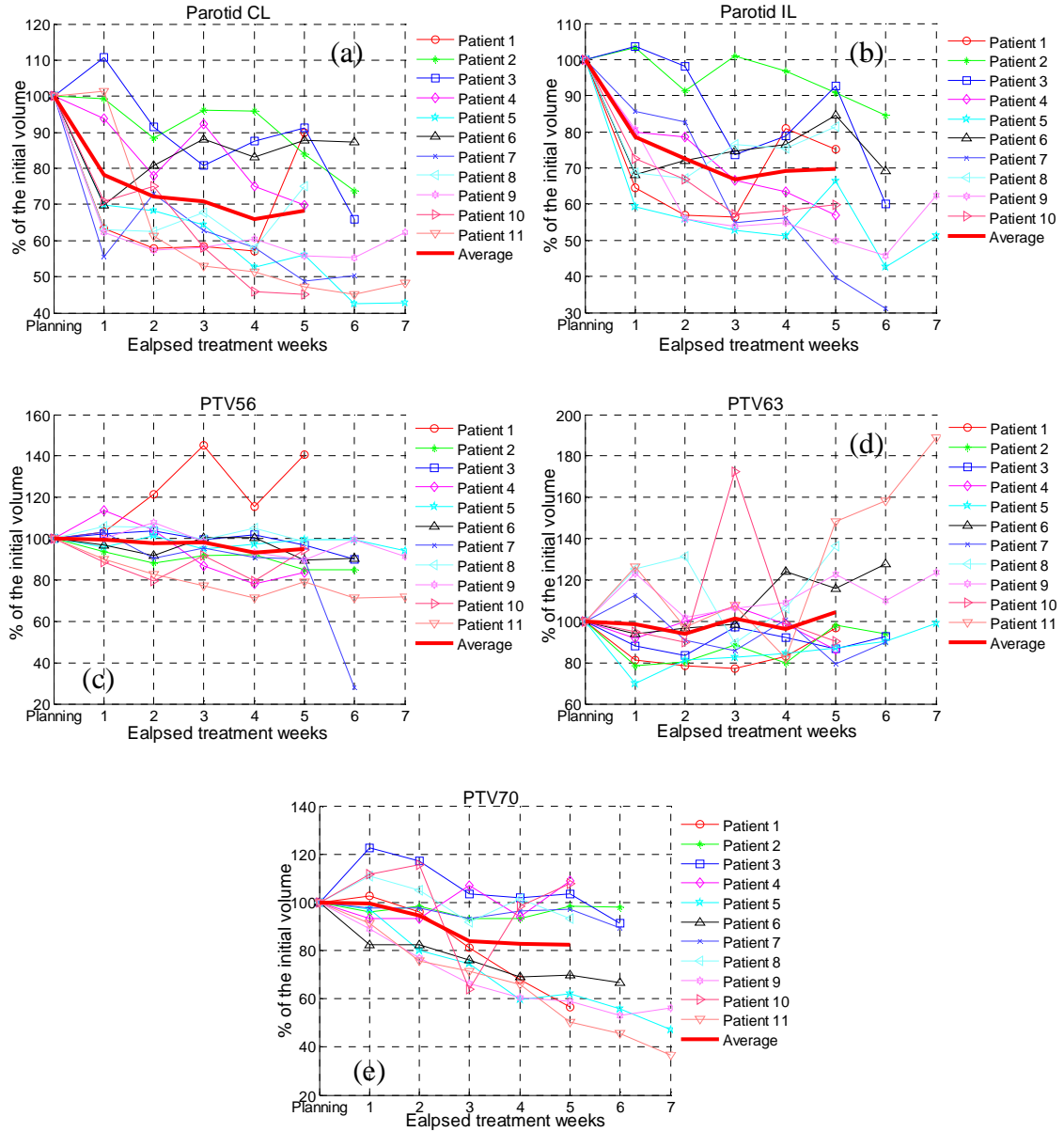
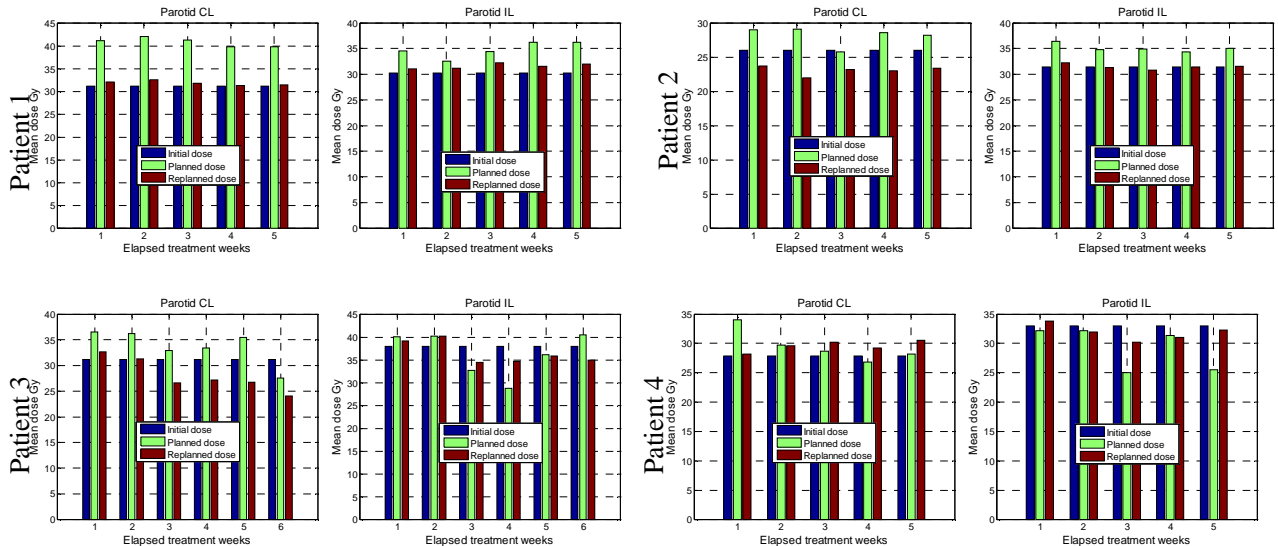


图 5-5 在放射治疗过程中, 体模内各器官的体积随放射治疗时间的变化趋势: (a) 异侧腮腺(CL); (b)同侧腮腺(IL); (c)PTV56Gy; (d) PTV63Gy; (e)PTV70Gy。



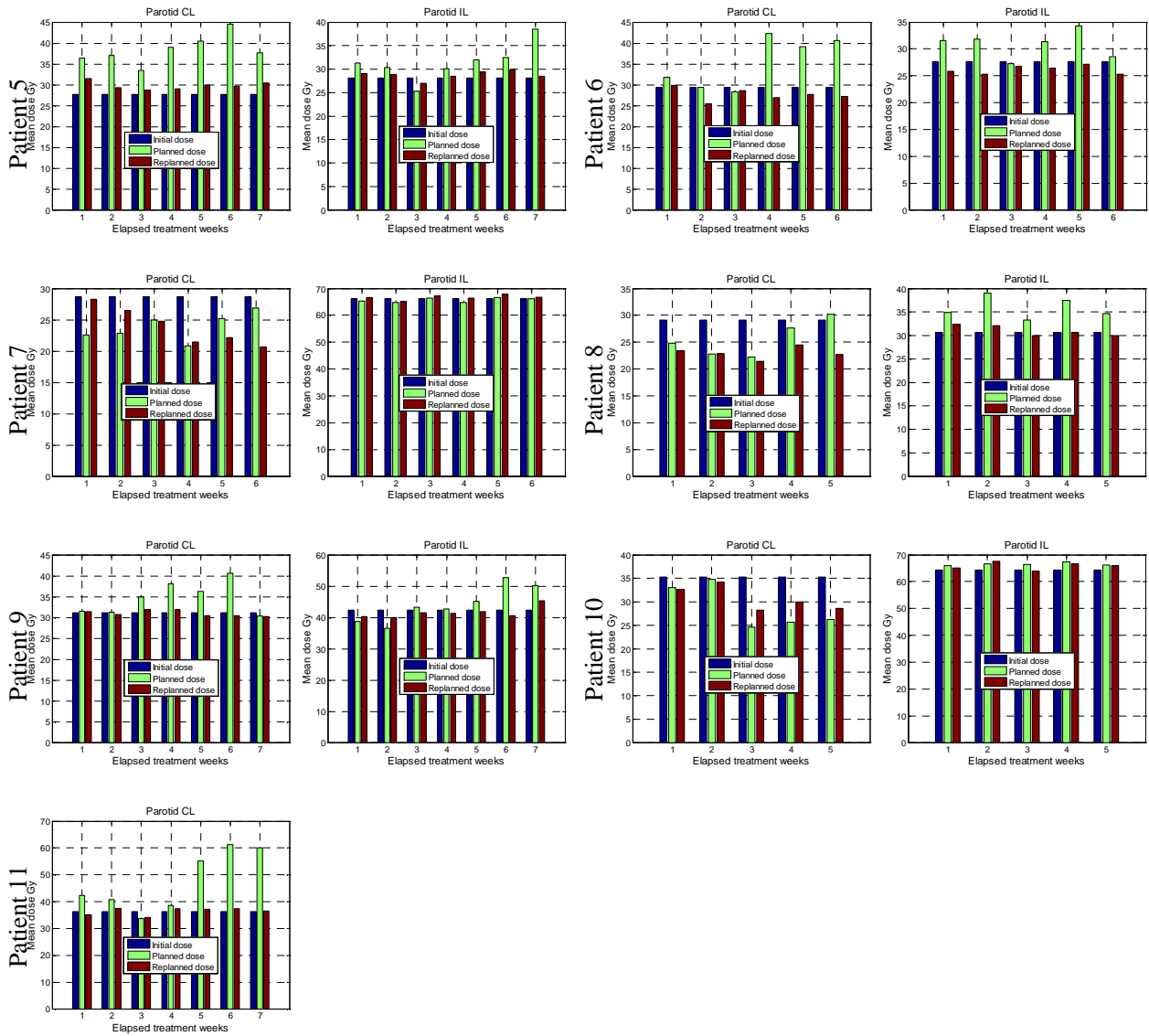
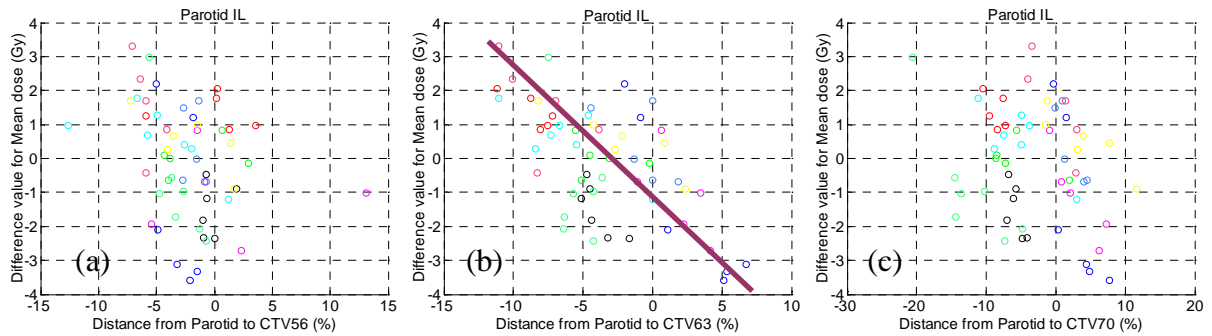


图 5-6 在放射治疗过程中, 11 例头颈部肿瘤患者分别在初始剂量分布(Initial dose)、初始计划在每周获取 CT 数据上的剂量分布(Planned dose)和每周重新计划的剂量分布(Replanned dose)中两腮腺组织的平均剂量值。图中, 从上到下, 左边两列分别为病人 1、3、5、7、9 和 11 的平均剂量分布, 右边两列分别为病人 2、4、6、8 和 10 的平均剂量分布。每个病人对应两幅图像, 左边的是异侧(CL)腮腺平均剂量分布, 右边是同侧(IL)腮腺剂量分布。



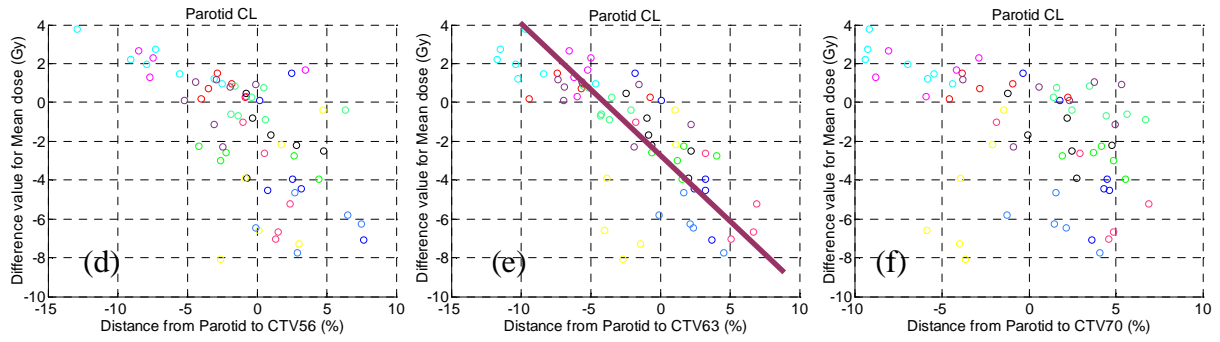
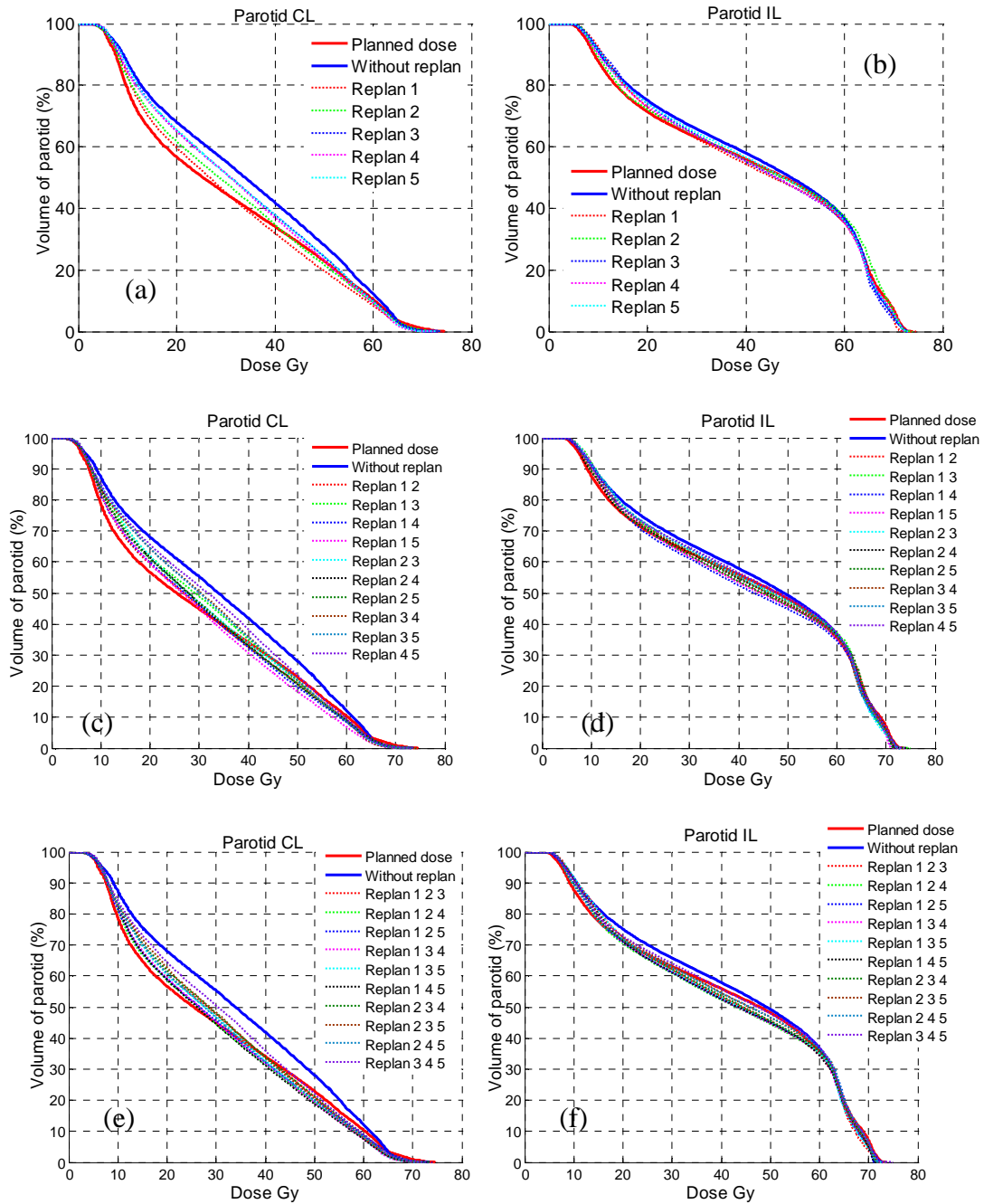


图 5-7 每周重新计划剂量分布(Replanned dose)的平均剂量与初始剂量分布(Initial dose)的平均剂量之差和腮腺中心到各靶区中心之间距离的关系。



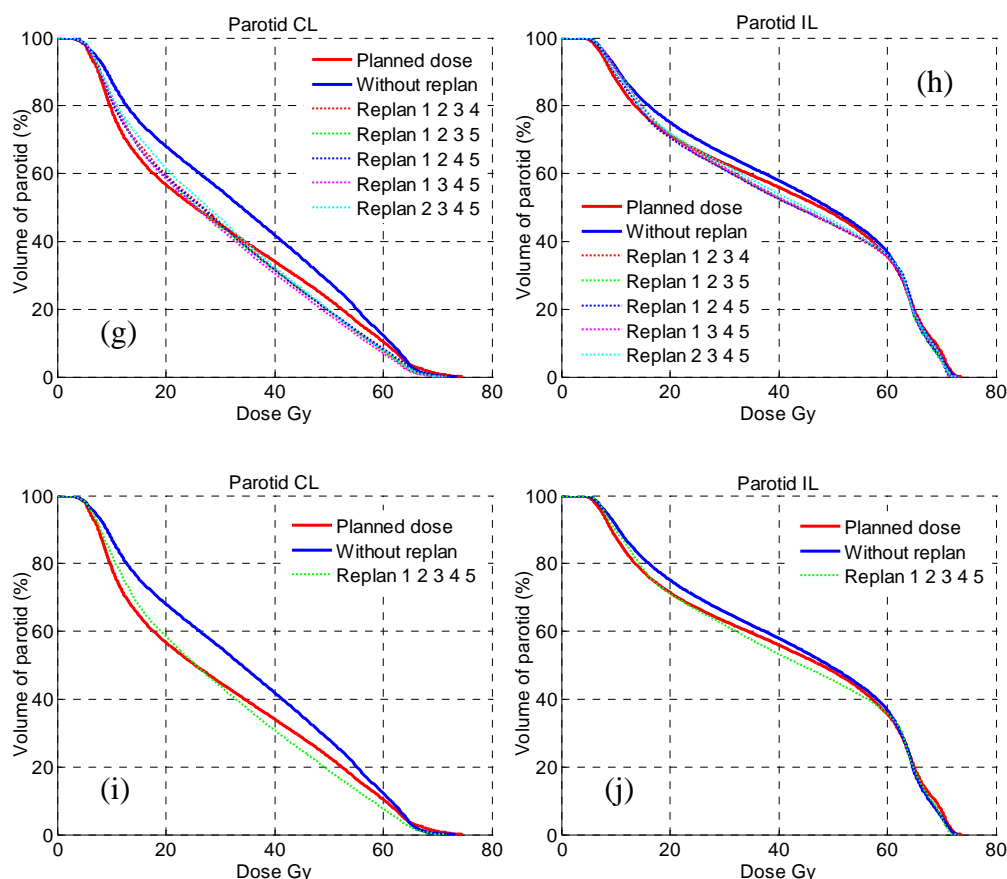


图 5-8 在不同重新计划次数下，不同重新计划时间组合的累积剂量分布的 DVH 曲线。图中从上到下分别为重新计划 1、2、3、4 和 5 次的剂量分布。左边一列为异侧腮腺的累积剂量分布，右边一列为同侧腮腺的累积剂量分布。

5.4.2.3 不同重新规划时间组合的累积剂量分布

已有的 11 例病人所需的治疗时间是不同的。病人 1、2、4、8 和 10 需要 5 周的治疗时间，病人 3、6 和 7 需要 6 周的治疗时间，而剩余病人则需要 7 周的治疗时间。为了能够得到这 11 例病人的平均累积剂量分布，我们只考虑病人 3、5、6、7、9 和 11 的前 5 次重新规划。如表 5.1 所示，所有不同重新规划次数下的不同重新规划时间的组合有 31 种。图 5-8 按不同的重新规划次数分别给出异侧腮腺组织和同侧腮腺组织内剂量分布的 DVH 曲线。同时给出初始计划的剂量分布(Planned dose)的 DVH 曲线和无重新规划的累积剂量分布(Without replan)的 DVH 曲线作为这 31 种累积剂量分布 DVH 曲线的参照物。从图中可以观察到，重新规划对异侧腮腺组织的剂量减少要明显多于同侧腮腺组织。这是因为异侧腮腺距离靶区较远，可以尽可能地把异侧腮腺从高剂量分布中分离出来；而同侧腮腺距离靶区近或在靶区内，在保证靶区剂量覆盖的同时，将同侧腮腺从高剂量区域中分离出来比较困难。我们进一步定量地分析上面提到的 33 种剂量分布，分别计算这些剂量分布中同侧腮腺和异侧腮腺组织的平均剂量值和 NTCP 值，如图 5-9 所示。从图中可以观察到，随着重新规划的次数增加，两腮腺组织内的平均剂量值低于初始剂量分布的平均值(图中红线)的概率越来越大。因此，随着重新规划的次数增加，重新规划对正常组织的保护作用也越来越明显。但是，并不是重新规划五次对正常组织的保护就是最好的，即平均剂量值和 NTCP 值最小。这是因此在重新规划的过程中，我们首先考虑靶区剂量的覆盖，然后才考虑减少正常组织上的剂量分布。为了综合地考虑重新规划对整体正常组织的影响，图 5-10 给出了在 33 种剂量分布中两腮腺组织的平均剂量值和 NTCP 值。由于两腮腺组织的贡献相同，因此其权重也相同。

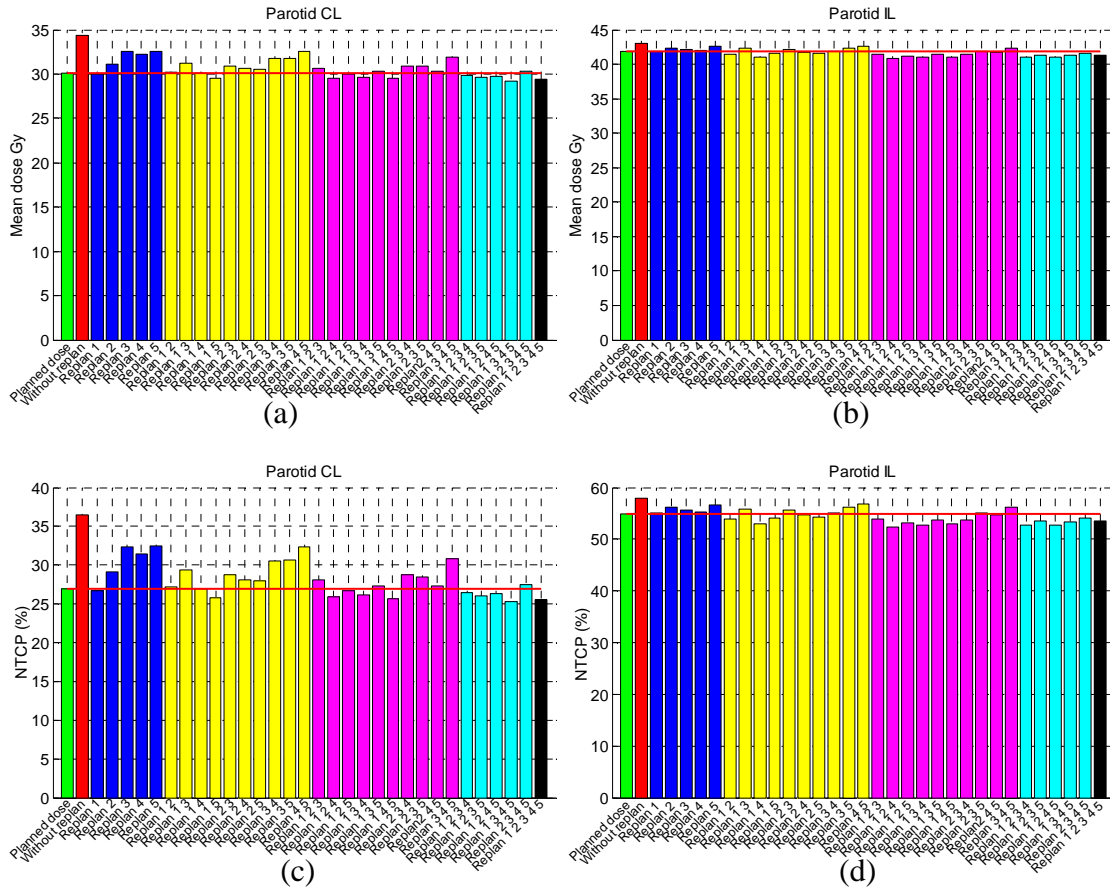


图 5-9 初始计划的剂量分布(Planned dose)、无重新计划的累积剂量分布(Without replan)和不同重新计划时间组合(31 种) 的剂量分布中,同侧腮腺和异侧腮腺组织的平均剂量值和 NTCP 值。(a)异侧腮腺的平均剂量值; (b) 同侧腮腺的平均剂量值; (c) 异侧腮腺的 NTCP 值; (d) 同侧腮腺的 NTCP 值。

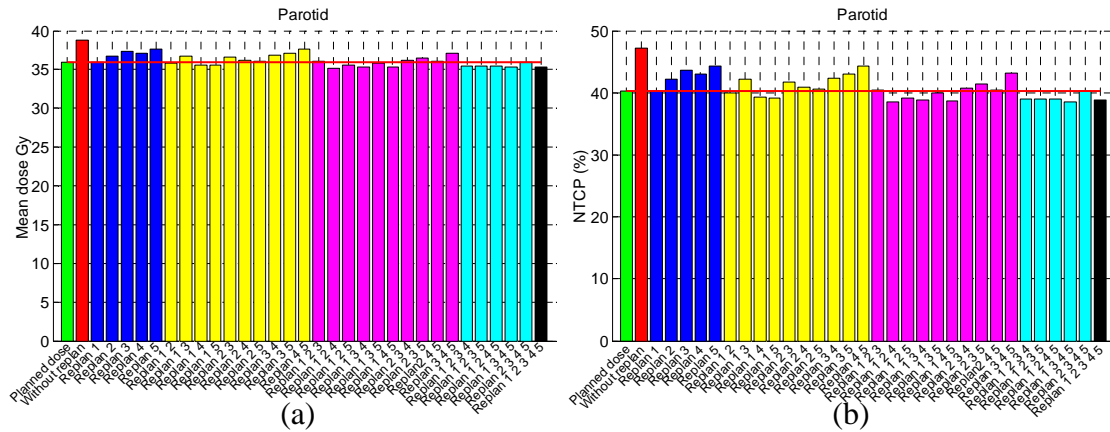


图 5-10 初始计划的剂量分布(Planned dose)、无重新规划的累积剂量分布(Without replan)和不同重新计划时间组合(31 种) 的剂量分布中,两腮腺组织的平均剂量值和 NTCP 值。(a)两腮腺的平均平均剂量值; (b)两腮腺的平均 NTCP 值。

5.4.2.3 最佳重新规划时间及次数

综上所述,我们已经计算得到所有不同重新规划次数下不同重新规划时间组合的腮腺组织 DVH 曲线以及平均剂量值和 NTCP 值。在此基础上,我们分两种方式分别获取最佳的组合,从而得到最佳的重新规划时间及次数。这两种选取最佳组合方式的主要区别在于是否考

考虑腮腺组织的过剂量照射(over-irradiated)。

在不考虑腮腺组织的过剂量照射情况下, 最佳组合是在前面 31 种组合中选择最小 EUD 值或 NTCP 值所对应的组合得到。表 5.2 为从图 5-9(c)和(d)中分别获取的同侧腮腺和异侧腮腺的最佳组合集, 即所有组合中 NTCP 值最小的 10%的所有组合。对于同侧腮腺, 最佳组合是“Replan 1 2 4”, 最佳的重新规划时间是第一、二、四周, 最佳重新规划次数是 3 次。与无重新规划的累积剂量分布(Without replan)相比, 该最优组合减少了同侧腮腺组织的剂量分布, EUD 值减少为 2.26Gy, 对应 NTCP 值减少 5.59%, 且在两种剂量分布下的同侧腮腺组织 DVH 曲线互不相关 ($p<0.05$)。对于异侧腮腺, 最佳组合是“Replan 1 3 4 5”, 即最佳重新规划时间为第一、三、四、五周, 最佳重新规划次数为 4 次。与无重新规划的累积剂量分布相比, 该最优组合也减少了异侧腮腺组织所接受到的剂量照射, EUD 值减少为 5.14Gy, 对应 NTCP 值减少 11.26%, 且异侧腮腺组织内的两 DVH 曲线互不相关($p<0.05$)。由此可见, 同侧腮腺组织与异侧腮腺组织的最佳重新规划时间及次数是不同的。我们综合地考虑前面 31 种组合对两腮腺组织的整体剂量改进, 得到最佳的全局最优组合。表 5.3 为从图 5-10 (b)中获取的两腮腺组织的最佳组合集。对于整体的腮腺组织, 最佳的组合是“Replan 1 2 4”, 最佳的全局重新规划时间是第一、二、四周, 最佳的全局重新规划次数为 3 次。最佳全局最优组合减少了正常组织的剂量分布, 平均 EUD 值减少 3.53Gy, 平均 NTCP 值减少 8.66%。

表 5.2 在不考虑腮腺组织的过剂量照射情况下, 同侧腮腺和异侧腮腺的最佳组合集

	最优组合	EUD (Gy)	NTCP	p 值
同侧腮腺	Replan 1 2 4	40.85	52.37%	1.49E-06
	Replan 1 3 4	41.02	52.79%	3.22E-05
	Replan 1 2 3 4	41.00	52.76%	2.73E-05
异侧腮腺	Replan 1 5	29.51	25.74%	0.0001
	Replan 1 2 4	29.58	25.89%	0.0046
	Replan 1 4 5	29.49	25.71%	0.0006
	Replan 1 3 4 5	29.24	25.21%	4.89E-06
	Replan 1 2 3 4 5	29.41	25.55%	0.0001

表 5.3 在不考虑腮腺组织的过剂量照射情况下, 两腮腺组织的最佳组合集

最优组合	EUD (Gy)	NTCP
Replan 1 2 4	35.21	38.45%
Replan 1 3 4	35.36	38.80%
Replan 1 4 5	35.29	38.62%
Replan 1 2 3 4	35.41	38.93%
Replan 1 2 4 5	35.41	38.92%
Replan 1 3 4 5	35.26	38.56%
Replan 1 2 3 4 5	35.35	38.78%

在考虑腮腺组织的过剂量照射情况下, 最佳组合是在前面 31 种组合中剂量分布高于初始计划剂量分布的组合(即图 5-9 和图 5-10 中 EUD 值或 NTCP 值高于红线的组合)中选择最小 EUD 值或 NTCP 值所对应的组合得到。表 5.4 为在这种情况下同侧腮腺和异侧腮腺的最佳组合。对于同侧腮腺, 有 38.71%(N=12)组合的剂量分布高于初始剂量分布。其中, 最优组合为“Replan 1”, 最佳重新规划时间为第一周, 最佳重新规划次数为 1 次。相比无重新规划的累积剂量分布, 最优组合减少了同侧腮腺的剂量分布, EUD 值减少了 1.19Gy, NTCP 减少了

2.93%，但两剂量分布的 DVH 曲线相关性比较大($p>0.05$)。对于异侧腮腺，有 51.61%($N=16$)组合的剂量分布高于初始剂量分布。其中，最优组合为“Replan 1 2”，最佳重新规划时间为第一、二周，最佳重新规划次数为 2 次。相比无重新规划的累积剂量分布，最优组合减少了异侧腮腺接受到的剂量分布，EUD 值减少了 4.17Gy，NTCP 减少了 9.29%，同样两剂量分布的 DVH 曲线相关性比较大($p>0.05$)。表 5.5 为考虑腮腺组织的过剂量照射情况下两腮腺组织的最佳组合。对于整体的腮腺组织，有 51.61%($N=16$)组合的剂量分布高于初始剂量分布。最佳的组合是“Replan 2 4 5”，最佳的全局重新规划时间是第二、四、五周，最佳的全局重新规划次数为 3 次。最佳全局最优组合减少了正常组织接收到的剂量分布，平均 EUD 值减少 2.68Gy，平均 NTCP 值减少 6.62%。

表 5.4 在考虑腮腺组织的过剂量照射情况下，同侧腮腺和异侧腮腺的最佳组合集

	最优组合	EUD (Gy)	NTCP	p 值
同侧腮腺	Replan 1	41.919	55.033%	0.5687
	Replan 3 4	41.920	55.035%	0.6894
	Replan 2 3 5	41.95	55.12%	0.6453
异侧腮腺	Replan 1 2	30.21	27.18%	0.4086
	Replan 1 3 5	30.29	27.34%	0.1783
	Replan 2 4 5	30.31	27.39%	0.2675

表 5.5 在考虑腮腺组织的过剂量照射情况下，两腮腺组织的最佳组合集

最优组合	EUD (Gy)	NTCP
Replan 2 5	36.10	40.59%
Replan 1 2 3	36.063	40.50%
Replan 2 4 5	36.060	40.49%

5.4.3 讨论

在这一节中，我们以头颈部肿瘤为例，提出了一种确定最佳重新规划时间及次数的方法。在新方法中，我们计算在不同重新规划时间及次数下病人体内的累积剂量分布，并根据累积剂量分布评价所有不同时间进行重新规划的组合对腮腺组织的保护作用。最佳的重新规划时间及次数为对腮腺组织保护最好的组合(即 EUD 值或 NTCP 值最小的组合)所对应的重新规划时间及次数。相对于直接根据体模内解剖结构变化或剂量变化确定最佳优化时间的方法^[3-6]，新方法综合考虑了重新规划在当前分次照射以及之后的分次照射中的剂量分布。从而全面地考虑了重新规划对整体剂量的影响，而不仅仅是某次分次照射最优。根据是否考虑腮腺组织的过剂量照射，我们给出了两种情况下的最佳重新规划时间及次数。在不考虑腮腺组织的过剂量照射的情况下，最佳的重新规划时间为第一、二、四周，重新规划 3 次。在考虑腮腺组织的过剂量照射的情况下，最佳的重新规划时间为二、四、五周，最佳重新规划次数为 3 次。

在本节中，我们统计了 11 例头颈部肿瘤患者同侧腮腺、异侧腮腺、PTV56Gy、PTV63Gy 和 PTV70Gy 五个器官的体积随治疗时间的变化趋势。与一些文献^[34-37]中的报道相同，平均腮腺体积随着放射治疗的进行不断缩小。但是，个体腮腺体积的变化与平均腮腺体积的变化差别还是很大的，如图 5-5 所示。因此，直接用平均腮腺体积来估算个体腮腺的体积可能会造成比较大的误差。在已有的文献^[34-37]中也发现了类似的结果，但这些文献过多地强调平均腮腺体积随治疗时间的变化趋势，而没有讨论平均腮腺体积的变化是否能代替个体病人腮腺体

积的变化, 及其适用范围。我们还尝试着寻找腮腺体积变化及位置变化对重新规划的影响。随着腮腺组织中心到靶区中心距离的增加, 重新规划后腮腺内的平均剂量与初始剂量分布的平均剂量的差值越小(如图 5-7 (b)和(c)所示), 重新规划对正常组织的保护作用越明显。在病人体内, 腮腺组织的位置靠近靶区 PTV63Gy。腮腺组织距离靶区 PTV63Gy 的距离越远, 越有利于重新规划将正常组织从靶区分离出来, 腮腺内的剂量分布越低。

最佳的重新规划时间及次数还与我们所考虑的器官有关。如前面介绍, 同侧腮腺和异侧腮腺的最佳重新规划时间及次数是不同的。在本节中, 我们只针对两个腮腺来确定最佳重新规划时间及次数。由于两个腮腺对人体的贡献是相同的, 其权重也应该是相同的(式 5.10)。因此, 我们根据两个腮腺的平均剂量值计算得到全局最佳重新规划时间及次数。在临床应用中, 还需要考虑腮腺之外的其它正常组织及靶区的剂量分布。为了进一步减少重新规划的次数, 我们放松了对腮腺组织剂量分布的限制, 在考虑腮腺组织过剂量照射的情况下, 确定最佳的重新规划次数及时间。实验表明, 在允许过剂量照射的情况下, 分别减少了同侧和异侧腮腺组织的重新规划次数, 如表 5.2 和表 5.4 所示。但是, 同时考虑整体的腮腺组织, 并没有减少重新规划的次数。

5.5 本章小结

在本章中, 我们首先介绍了自适应放射治疗技术的组成部分及累加剂量的计算方法。其中, 主要介绍了离线 ART 方法, 并提出了总剂量分布的概念, 从而将通常使用物理准则的离线 ART 方法扩展到能够使用所有的准则。然后, 我们以头颈部肿瘤为例, 提出了一种确定离线 ART 中最佳重新规划时间及次数的方法。新方法根据累积剂量分布来确定最佳重新规划时间及次数, 综合地考虑了重新规划对整体剂量的影响, 而不仅仅是对某次分次照射进行校正。最后, 根据是否考虑腮腺组织的过剂量照射, 给出了两种情况下的最佳重新规划的时间及次数。

5.6 参考文献

- [1] Jacob R, Hanlon AL, Horwitz EM, et al. The relationship of increasing radiotherapy dose to reduced distant metastases and mortality in men with prostate cancer [J]. *Cancer*, 2004, 100(3):538-543.
- [2] Wang JZ, Li XA. Impact of tumor repopulation on radiotherapy planning [J]. *Int J Radiat Oncol Biol Phys*, 2005, 61(1):220-227.
- [3] Chen J, Morin O, Aubin M, et al. Dose-guided radiation therapy with megavoltage cone-beam CT [J]. *Br J Radiol*, 2006, 1:S87-S98.
- [4] Ahamad A, Dong L, Zhang L, et al. Is there a trigger point for adaptive replanning during head & neck IMRT [J]? *Int J Radiat Oncol Biol Phys*, 2006, 66(3):S100-101. (Abstract)
- [5] Meldolesi E, Wu Q, Chen P, et al. Evaluation of anatomic and dosimetric changes during treatment course of head and neck (HN) IMRT: Is replanning necessary [J]? *Int J Radiat Oncol Biol Phys*, 2006, 66(3):S101-102. (Abstract)
- [6] Siddiqui F, Boyle D, Weldon M, et al. Adaptive radiation therapy in head and neck Cancers: When should we re-plan [J]? *Int J Radiat Oncol Biol Phys*, 2011, 81(2):S494-495. (Abstract)
- [7] Wu Q, Nandalur SR, Nandalur KR, et al. What time is optimal for replanning head and neck IMRT (HN-IMRT) [J]? *Int J Radiat Oncol Biol Phys*, 2008, 72(1):S163-163. (Abstract)
- [8] Wu QJ, Li T, Wu Q, et al. Adaptive radiation therapy: technical components and clinical applications [J]. *Cancer J*, 2011, 17(3):182-189.
- [9] Wu QJ, Thongphiew D, Wang Z, et al. On-line re-optimization of prostate IMRT plans for adaptive radiation therapy [J]. *Phys Med Biol*, 2008, 53(3):673-691.
- [10] Yan D, Jaffray DA, Wong JW. A model to accumulate fractionated dose in a deforming organ

- [J]. *Int J Radiat Oncol Biol Phys*, 1999, 44(3):665-675.
- [11] Coselmon MM, Balter JM, McShan DL, et al. Mutual information based CT registration of the lung at exhale and inhale breathing states using thin-plate splines [J]. *Med Phys*, 2004, 31(11):2942-2948.
- [12] Lu W, Olivera GH, Chen Q, et al. Deformable registration of the planning image (kVCT) and the daily images (MVCT) for adaptive radiation therapy [J]. *Phys Med Biol*, 2006, 51(17):4357-4374.
- [13] Kashani R, Hub M, Balter JM, et al. Objective assessment of deformable image registration in radiotherapy: a multi-institution study [J]. *Med Phys*, 2008, 35(12):5944-5953.
- [14] Mayr NA, Wang JZ, Zhang D, et al. Longitudinal changes in tumor perfusion pattern during the radiation therapy course and its clinical impact in cervical cancer [J]. *Int J Radiat Oncol Biol Phys*, 2010, 77(2):502-508.
- [15] Roper J, Bowsher J, Yin FF. On-board SPECT for localizing functional targets: a simulation study [J]. *Med Phys*, 2009, 36(5):1727-1735.
- [16] Chung PW, Haycocks T, Brown T, et al. On-line aSi portal imaging of implanted fiducial markers for the reduction of interfraction error during conformal radiotherapy of prostate carcinoma [J]. *Int J Radiat Oncol Biol Phys*. 2004, 60(1):329-334.
- [17] Jaffray DA, Siewerdsen JH, Wong JW, et al. Flat-panel cone-beam computed tomography for image-guided radiation therapy [J]. *Int J Radiat Oncol Biol Phys*, 2002, 53(5):1337-1349.
- [18] Wu Q, Ivaldi G, Liang J, et al. Geometric and dosimetric evaluations of an online image-guidance strategy for 3D-CRT of prostate cancer [J]. *Int J Radiat Oncol Biol Phys*, 2006, 64(5):1596-1609.
- [19] Yan D, Vicini F, Wong J, et al. Adaptive radiation therapy [J]. *Phys Med Biol*, 1997, 42(1):123-132.
- [20] Wu C, Jeraj R, Olivera GH, et al. Re-optimization in adaptive radiotherapy [J]. *Phys Med Biol*. 2002, 47(17):3181-3195.
- [21] de la Zerda A, Armbruster B, Xing L. Formulating adaptive radiation therapy (ART) treatment planning into a closed-loop control framework [J]. *Phys Med Biol*, 2007, 52(14):4137-4153.
- [22] Saka B, Rardin RL, Langer MP. Adaptive intensity modulated radiation therapy planning optimization with changing tumor geometry and fraction size limits [J]. *IEEE Transactions on Healthcare Systems Engineering*, 2011, 1(4):247-263.
- [23] Sir MY, Epelman MA, Pollock SM. Stochastic programming for off-line adaptive radiotherapy [J]. *Ann Oper Res*, 2012, 196(1):767-797.
- [24] Yan D, Liang J. Expected treatment dose construction and adaptive inverse planning optimization: implementation for offline head and neck cancer adaptive radiotherapy [J]. *Med Phys*, 2013, 40(2): 021719-10.
- [25] Mohan R, Zhang X, Wang H, et al. Use of deformed intensity distributions for on-line modification of image-guided IMRT to account for interfractional anatomic changes [J]. *Int J Radiat Oncol Biol Phys*, 2005, 61(4):1258-1266.
- [26] Court LE, Tishler RB, Petit J, et al. Automatic online adaptive radiation therapy techniques for targets with significant shape change: a feasibility study [J]. *Phys Med Biol*, 2006, 51(10):2493-2501.
- [27] Feng Y, Castro-Pareja C, Shekhar R, et al. Direct aperture deformation: an interfraction image guidance strategy [J]. *Med Phys*, 2006, 33(12):4490-4498.
- [28] Ahunbay EE, Peng C, Godley A, et al. An on-line replanning method for head and neck adaptive radiotherapy [J]. *Med Phys*, 2009, 36(10):4776-4790.

- [29] Li T, Thongphiew D, Zhu X, et al. Adaptive prostate IGRT combining online re-optimization and re-positioning: a feasibility study [J]. *Phys Med Biol*, 2011, 56(5):1243-1258.
- [30] Cachier P, Bardinet E, Dormont D, et al. Iconic feature based nonrigid registration: the PASHA algorithm [J]. *Comput Vis Image Underst*, 2003, 89(23):272–298.
- [31] Vercauteren T, Pennec X, Perchant A, et al. Symmetric log-domain diffeomorphic Registration: a demons-based approach [J]. *Med Image Comput Comput Assist Interv*, 2008, 11(Pt 1):754-761.
- [32] Dijkema T, Raaijmakers CP, Ten Haken RK, et al. Parotid gland function after radiotherapy: the combined michigan and utrecht experience [J]. *Int J Radiat Oncol Biol Phys*, 2010, 78(2):449-453.
- [33] Burman C, Kutcher GJ, Emami B, et al. Fitting of normal tissue tolerance data to an analytic function [J]. *Int J Radiat Oncol Biol Phys*, 1991, 21(1):123-135.
- [34] Barker JL Jr, Garden AS, Ang KK, et al. Quantification of volumetric and geometric changes occurring during fractionated radiotherapy for head-and-neck cancer using an integrated CT/linear accelerator system [J]. *Int J Radiat Oncol Biol Phys*, 2004, 59(4):960-970.
- [35] Wu Q, Chi Y, Chen PY, et al. Adaptive replanning strategies accounting for shrinkage in head and neck IMRT [J]. *Int J Radiat Oncol Biol Phys*, 2009, 75(3):924-932.
- [36] Castadot P, Lee JA, Geets X, et al. Adaptive radiotherapy of head and neck cancer [J]. *Semin Radiat Oncol*, 2010, 20(2):84-93.
- [37] Schwartz DL, Dong L. Adaptive radiation therapy for head and neck cancer-can an old goal evolve into a new standard [J]? *J Oncol*, 2011, 2011: 690595-690608.

第六章 总结和展望

6.1 全文工作总结

在设计肿瘤放射治疗方案过程中,准确地预测剂量分布是放射治疗质量保证的关键环节。在对诊断和治疗信息整合的基础上,放射治疗系统模拟整个治疗过程并制定出最佳的放射治疗方案,以保证在靶区肿瘤细胞得到足够的致死剂量的同时,周围正常组织所接收的照射量尽可能的少。一个完整的、高性能的放射治疗系统需要大量关键技术的支持如:医学图像的快速三维重建、高精度的三维剂量计算、治疗计划的方案优化或逆向设计、功能强大的放射治疗实施设备等。因此,针对精确放射治疗剂量计算方法及方案优化方法的研究是一项非常有意义的研究课题,并吸引了众多学者的关注。本文主要致力于笔形束剂量计算方法和方案优化算法的研究,在总结现有笔形束剂量计算方法及方案优化算法的基础上,对已有的经典笔形束算法进行改进,增加剂量计算的精度;将生物准则包含到方案优化的目标函数中,保证方案优化结果为全局生物最优。在前面研究的基础上,我们进一步研究了自适应放射治疗中最佳的重新规划时间及次数和随机几何不确定性误差的校正。概括全文的研究成果和贡献,主要有如下几个方面:

6.1.1 改进的笔形束剂量计算方法

1. 介绍笔形束剂量计算方法及基于核模型剂量计算方法的一些基本假设,如射线源几何结构及能谱变化、非均匀介质内剂量分布的校正、剂量卷积/叠加的方法。
2. 详细地介绍改进的坐标系统,将在用蒙特卡罗软件获取的笔形束核数据转换到改进的坐标系统中。笔形束核分为深度方向分量和侧向分量两部分。分别提取并存储深度方向分量和侧向分量的参数值。
3. 在球壳坐标系统中,经典算法以射线与体模表层的交点为参考点,进行非均匀介质中的剂量校正。在这种校正方法中,剖面分量校正采用一种近似计算的方法,降低了剂量计算的精度。而且剖面分量中包含了深度信息,不能利用快速变换方法对其进行加速。针对经典算法中的非均匀剂量校正方法,我们提出以射线与球壳的碰撞点为参考点进行非均匀介质中的剂量校正。新方法一方面避免了在剖面分量校正时所使用的近似计算,另一方面将深度信息从剖面分量中分离出来。在不考虑剖面非均匀介质分布的情况下,可以在每一层球壳上直接使用 FFTC 技术对新提出的剂量计算方法进行加速。
4. 在不同的假设体模上验证新的笔形束剂量计算模型的有效性。以蒙特卡罗算法的计算结果为金标准,并与 AAA 算法的计算结果进行比较。与 AAA 算法相比,新方法在增加剂量计算精度的同时,没有增加剂量计算所需的时间。利用 FFTC 技术对新算法进行加速,计算速度提升了将近 40 倍。

6.1.2 基于生物准则的 IMRT 方案优化算法

1. 总结了通常被用来评估放射治疗质量的物理准则和生物准则。其中物理准则包含最小剂量准则、最大剂量准则、平均剂量准则和 DVH 准则。生物准则主要包含 EUD 准则、TCP 准则和 NTCP 准则。其中,全部的物理准则和部分生物准则已在方案优化中使用,用来控制体模内各器官的剂量分布。
2. 详细地介绍了使用物理准则的方案优化算法的缺点,及使用生物准则的方案优化算法的优点。
3. 直接将非凸的准则作为目标函数和约束条件会导致优化问题变为非凸优化问题,增加了求解优化问题的难度。在总结将生物准则转换成等效凸生物准则方法的基础上,提出了利用等效凸生物准则构造目标函数的方法,并给出基于等效凸生物准则的子目标函数及其一阶导数。
4. 以前列腺肿瘤患者为例,验证了基于等效凸 NTCP 准则的 IMRT 方案优化算法的有效性。利用基于梯度的优化算法, L-BFGS, 求解 FMO 问题,并与基于 DVH 准则的方案优化算法

在 10 例前列腺肿瘤患者上做了对比实验。结果表明, 利用新算法可以得到临床可接受的方案优化结果。基于 NTCP 准则的方案优化算法能够直接估算危及器官出现并发症的概率, 与临床治疗密切相关, 因此新算法能够提供对整体治疗效果最有利的方案优化结果。新算法还与其它两种基于 NTCP 准则的方案优化算法进行了比较。对比实验表明, 新方法可以得到更高质量的治疗计划, 同时其计算所需的时间也要少于其它两种基于 NTCP 准则的方案优化算法。

6.1.3 IMRT 中随机几何不确定性的校正

1. 首先, 详细地介绍了 IMRT 中的几何不确定性的来源, 以及描述随机几何不确定性的概率密度函数及其参数取值。然后, 总结了校正随机几何不确定性的序列反卷积方法。
2. 通过在频域对已有的序列反卷积方法进行分析可知, 反卷积核算法优于直接序列反卷积算法是因为反卷积核算法等于直接序列展开算法和一个低通滤波器的乘积, 可以进一步抑制反卷积结果中的高频信息。由于这个低通滤波器的性能与序列展开的截止项有关, 其输出性能不能随便修改。因此, 我们使用经典的滤波器进行滤波, 并根据随机几何不确定性的运动幅度修改滤波参数, 使得卷积后的注量图与静态注量图的差值尽可能的小。
3. 我们在两个 2D 注量图和一个前列腺肿瘤病例上验证了新算法的有效性。实验结果表明, 利用巴特沃斯滤波器可以很好地滤除直接序列展开结果中的高频成份, 减少卷积后注量图中的冷点和热点。在注量图内的平滑区域, 新算法可以取得和其它反卷积算法相同的效果。在射野的边缘位置, 新算法明显地减少了卷积后注量图与静态注量图的差异。

6.1.4 自适应放射治疗中的最佳重新规划时间及次数

1. 详细地介绍了自适应放射治疗技术的三个主要组成部分: 图像引导、计划调整和剂量验证。现有的离线 ART 技术主要是利用物理准则构造目标函数。我们提出了总剂量分布的概念, 使得离线 ART 方法能够使用所有的准则。
2. 在介绍计算累加剂量算法的基础上, 我们以头颈部肿瘤治疗为例, 提出了一种确定离线 ART 中最佳重新规划时间及次数的方法。重新规划需要大量的资源, 由于医疗资源有限, 不可能为每个病人在每次分次照射之前进行重新规划。但是, 可以使用有限次数的重新规划, 最大程度地改善病人的放射治疗质量。相对于直接根据体模内解剖结构或剂量变化确定最佳优化时间的方法, 新方法根据累积剂量分布来确定最佳重新规划时间及次数, 综合地考虑了重新规划对整体剂量分布的影响, 而不仅仅是对某次分次照射进行校正。根据是否考虑腮腺组织的过剂量照射, 分别给出了两种情况下头颈部肿瘤自适应放射治疗的最佳重新规划时间及次数。

6.2 未来工作展望

本文主要针对精确放射治疗剂量计算及方案优化算法的部分关键问题展开了研究, 由于剂量计算及方案优化算法涉及了诸多学科, 还有不少问题有待解决, 未来的研究方向包括但不限于:

1. 剂量计算方法。本文第二章中, 在总结现有笔形束模型的基础上, 提出了一种新的笔形束剂量计算方法。在笔形束剂量计算过程中, 首先假设粒子沿深度方向传输, 然后再沿剖面方向到达剂量沉积点。然而, 实际中的射束能量并不是按这两个方向传输的。由于笔形束模型自身的局限性, 很难进一步提升笔形束剂量计算方法的精度。在对蒙特卡罗代码 BEAMnrc 和 Geant 4 进行研究的基础上, 在未来的研究中我们将使用 GPU 对蒙特卡罗算法进行加速, 以实现蒙特卡罗剂量计算方法在临床实践上的应用。
2. 方案优化方法。在本文第三章中, 我们将已有的等效凸生物准则应用到 IMRT 方案优化的目标函数中。在未来的研究中, 一方面为没有等效凸函数的生物准则找到等效凸变换。另一方面, 已有的方案优化算法是解决 FMO 问题, 为了进一步提高方案优化的精度, 我们将把基于等效凸生物准则的 FMO 方法扩展到直接子野优化方法上。
3. 自适应放射治疗。在本文第四章中, 我们研究了自适应放射治疗中的最佳重新规划的时间

及次数。在未来的研究中，我们主要针对在线 ART 和离线 ART 方法进行改进。在在线 ART 方面，主要是改进算法，加快修改计划的速度；在离线 ART 方面，完善离线 ART 方案优化可以使用的优化准则，进一步提高方案优化的质量。

致 谢

个人简历

姓 名:

性 别: 男

出生年月: 1984. 05. 01

籍 贯: 内蒙古

政治面貌: 团员

专 业: 计算机科学与技术

研究方向: 图像处理与科学可视化

教育经历:

工作经历: 无

攻读博士学位期间发表论文清单

1. 张鹏程, 朱健, 李元金, 舒华忠.一种改进的多叶准直器射野分割算法[J].东南大学学报: 自然科学版, 2012, 42(5):875-879.
doi: 10.3969/j.issn.1001-0505.2012.05.015 (第一作者 **EI** 索引)

2. Pengcheng Zhang, Renaud De Crevoisier, Antoine Simon, Pascal Haigron, Jean-Louis Coatrieux, Baosheng Li and Huazhong Shu. A new deconvolution approach to robust fluence for intensity modulation under geometrical uncertainty. Phys Med Biol, 2013, 58(17):6095-6110.
doi: 10.1088/0031-9155/58/17/6095 (第一作者 **SCI** 索引 **IF:2.7**)

3. Pengcheng Zhang, Huazhong Shu, Antoine Simon, Pascal Haigron and Renaud De Crevoisier. A new pencil beam model for photon dose calculations. 2013 6rd International Conference on Biomedical Engineering and Informatics (BMEI 2013) (第一作者 **EI** 索引)

4. Pengcheng Zhang, Antoine Simon, Huazhong Shua, Pascal Haigron, Caroline Lafond, Jean-Louis Coatrieux, Jian Zhu, Baosheng Li and Renaud De Crevoisier. IMRT optimization incorporating the equivalent convex NTCP constraint. (第一作者, 在审)

5. Pengcheng Zhang, Antoine Simon, Renaud De Crevoisier, Pascal Haigron, Mohamed Hatem Nassef, Baosheng Li and Huazhong Shu. A new pencil beam model for photon dose calculations in heterogeneous media. (第一作者, 在审)

VU :

Le Directeur de Thèse
(Nom et Prénom)

VU :

Le Responsable de l'École Doctorale

VU pour autorisation de soutenance

Rennes, le

Le Président de l'Université de Rennes 1

Guy CATHELINEAU

VU après soutenance pour autorisation de publication :

Le Président de Jury,
(Nom et Prénom)